Nucleic acid markers for use in determining predisposition to neoplasm and/or adenoma

FIELD OF THE INVENTION

The present invention relates generally to novel nucleic acid molecules, the levels and/or patterns of expression of which are indicative of the onset, predisposition to the onset and/or progression of a neoplasm and to derivatives, homologues or analogues of said molecules. More particularly, the present invention is directed to novel nucleic acid molecules, the levels of expression of which are indicative of the onset and/or progression of a gastrointestinal tract neoplasm, such as an adenoma, and to derivatives, homologues or analogues of said molecules. The present invention is further directed to isolated proteins encoded thereby and to derivatives, homologues, analogues, chemical equivalents and mimetics thereof. The molecules of the present invention are useful in a range of prophylactic, therapeutic and/or diagnostic applications including, but not limited to, those relating to the diagnosis and/or treatment of colorectal neoplasms such as colorectal adenomas. In a related aspect, the present invention is directed to a method of screening a subject for the onset, predisposition to the onset and/or progression of a neoplasm by screening for modulation in the level of expression of one or more nucleic acid molecule markers.

BACKGROUND OF THE INVENTION

Bibliographic details of the publications referred to by author in this specification are collected alphabetically at the end of the description.

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgment or any form of suggestion that that prior art forms part of the common general knowledge in Australia.

Adenomas are benign tumours of epithelial origin which are derived from glandular tissue or exhibit clearly defined glandular structures. Some adenomas show recognisable tissue

elements, such as fibrous tissue (fibroadenomas), while others, such as bronchial adenomas, produce active compounds giving rise to clinical syndromes. Tumours in certain organs, including the pituitary gland, are often classified by their histological staining affinities, for example eosinophil, basophil and chromophobe adenomas.

Adenomas may become carcinogenic and are then termed adenocarcinomas. Accordingly, adenocarcinomas are defined as malignant epithelial tumours arising from glandular structures, which are constituent parts of most organs of the body. This term is also applied to tumours showing a glandular growth pattern. These tumours may be sub-classified according to the substances that they produce, for example mucus secreting and serous adenocarcinomas, or to the microscopic arrangement of their cells into patterns, for example papillary and follicular adenocarcinomas. These carcinomas may be solid or cystic (cystadenocarcinomas). Each organ may produce tumours showing a variety of histological types, for example the ovary may produce both muconous and cystadenocarcinoma. In general, the overall incidence of carcinoma within an adenoma is approximately 5%. However, this is related to size and although it is rare in adenomas of less than 1 centimetre, it is estimated at 40 to 50% villous lesions which are greater than 4 centimetres. Adenomas with higher degrees of dysplasia have a higher incidence of carcinoma. Once a sporadic adenoma has developed, the chance of a new adenoma occurring is approximately 30% within 26 months.

Colorectal adenomas represent a class of adenomas which are exhibiting an increasing incidence, particularly in more affluent countries. The causes of adenoma, and its shift to adenocarcinoma, are still the subject of intensive research. To date it has been speculated that in addition to genetic predisposition, environmental factors (such as diet) play a role in the development of this condition. Most studies indicate that the relevant environmental factors relate to high dietary fat, low fibre and high refined carbohydrates.

Colonic adenomas are localised proliferations of dysplastic epithelium which are initially flat, but with increased growth from the mucosal forming adenomas. They are classified by their gross appearance as either sessile (flat) or penduculated (having a stalk). While

small adenomas (less than 0.5 millimetres) exhibit a smooth tan surface, penduculated adenomas have a head with a cobblestone or lobulated red-brown surface. Sessile adenomas exhibit a more delicate villous surface. Penduculated adenomas are more likely to be tubular or tubulovillous while sessile lesions are more likely to be villous. Sessile adenomas are most common in the cecum and rectum while overall penduculated adenomas are equally split between the sigmoid-rectum and the remainder of the colon.

The etiology of adenoma of the colon, and in particular the dysplasia-adenoma-carcinoma sequence is thought to occur in the setting of increasing loss of heterozygosity in genes involved in DNA replication accuracy, tumour suppression and oncogene activation. A hereditary predisposition to cancer is found in 1% of colorectal carcinoma patients and in 5-10% of patients with Hereditary Non-Adenomatosis Polyposis. It is thought that for each lesion the loss of heterozygosity must occur in multiple genes. Currently there are a number of mechanisms proposed to account for the known environment, dietary and genetic predispositions to colorectal cancer. Although no consensus has yet been reached, loss of heterozygosity appears to be a common feature.

Adenomas are generally asymptomatic, therefore rendering difficult their early diagnosis and treatment. It is technically impossible to predict the presence or absence of carcinoma based on the gross appearance of adenomas, although larger adenomas are thought to exhibit a higher incidence of concurrent malignancy than smaller adenomas. Sessile adenomas exhibit a higher incidence of malignancy than penduculated adenomas of the same size. Some adenomas result in the production of microscopic stool blood loss. However, since stool blood can also be indicative of non-adenomatous conditions and obstructive symptoms are generally not observed in the absence of malignant change, the accurate diagnosis of adenoma is rendered difficult without the application of highly invasive procedures such as biopsy analysis. Accordingly, there is an on-going need to elucidate not only the causes of adenoma and its shift to malignancy but to develop more informative diagnostic protocols, in particular protocols which will enable the rapid, routine and accurate diagnosis of adenoma at an early stage, such as the pre-malignant stage.

To date, research has focused on the identification of gene mutations which lead to the development of adenoma. In work leading up to the present invention, however, the inventors have surprisingly determined that changes in the level of expression of unmutated genes which are also expressed in healthy individuals are indicative of adenoma development. The inventors have further determined that in relation to colorectal adenomas, diagnosis can be made based on screening for the expression of mRNA gene transcripts corresponding to any one or more of the panel of genes disclosed herein. In this regard, the inventors have still further determined that some of the genes identified herein as being expressed in healthy individuals at significantly lower levels than that observed in individuals who have developed an adenoma do not correlate with any known gene sequences. Accordingly, the inventors have identified a panel of genes which, in addition to facilitating the diagnosis of adenoma development, further facilitate the development of prophylactic and therapeutic protocols directed to modulation of their expression and functional activity and thereby the development of therapeutic and/or prophylactic protocols for treating patients at risk of or who have developed adenomas.

SUMMARY OF THE INVENTION

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

The subject specification contains nucleotide sequence information prepared using the programme PatentIn Version 3.0, presented herein after the bibliography. Each nucleotide sequence is identified in the sequence listing by the numeric indicator <201> followed by the sequence identifier (eg. <210>1, <210>2, etc). The length, type of sequence (DNA, etc) and source organism for each nucleotide sequence is indicated by information provided in the numeric indicator fields <211>, <212> and <213>, respectively. Nucleotide sequences referred to in the specification are identified by the indicator SEQ ID NO: followed by the sequence identifier (eg. SEQ ID NO:1, SEQ ID NO:2, etc.). The sequence identifier referred to in the specification correlates to the information provided in numeric indicator field <400> in the sequence listing, which is followed by the sequence identifier (eg. <400>1, <400>2, etc). That is SEQ ID NO:1 as detailed in the specification correlates to the sequence indicated as <400>1 in the sequence listing.

A summary of the sequences detailed in this specification is provided prior to the examples.

One aspect of the present invention provides a method for determining the onset or a predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 1-2, SEQ ID NOs: 4-6, SEQ ID NOs: 8-32, SEQ ID NOs: 35-37 or SEQ ID NO: 59 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

In another aspect of the present invention there is provided a method for determining the onset or a predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NO: 38, SEQ ID NOs: 40-43, SEQ ID NOs: 45-49, SEQ ID NOs: 58-60, SEQ ID NO: 62, SEQ ID NOs: 64-66, SEQ ID NOs: 68-72 or SEQ ID NOs: 337-338 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

In yet another aspect of the present invention there is provided a method for determining the onset or a predisposition to the onset of a neoplasm in an individual, said method

comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 73-219 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

In still another aspect of the present invention there is provided a method for determining the onset or a predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 220-336 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

Another aspect of the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 1-2, SEQ ID NOs: 4-6, SEQ ID NOs: 8-32, SEQ ID NOs: 35-37 or SEQ ID NO: 59 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

In another aspect, the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NO: 38, SEQ ID NOs: 40-43, SEQ ID NOs: 45-49, SEQ ID NOs: 51-56, SEQ ID NOs: 58-60, SEQ ID NO: 62, SEQ ID NOs: 64-66, SEQ ID NOs: 68-72 or SEQ ID NOs: 337-338 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

In yet another aspect the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 73-219 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

In still another aspect the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 220-336 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or

a functional derivative, variant or homologue of said nucleic acid molecule

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

In a preferred embodiment, the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any three:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NO: 7, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 24, SEQ ID NO: 65, SEQ ID NO: 19, SEQ ID NO: 1, SEQ ID NO: 53, SEQ ID NO: 72, SEQ ID NO: 11 or SEQ ID NO: 26 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any three of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule,

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

In another preferred embodiment the present invention provides a method for determining the onset or the predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any four:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 4-6, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NOs: 21-22, SEQ ID NOs: 27-29, SEQ ID NOs: 30-31, SEQ ID NO: 36,

SEQ ID NOs: 37-38, SEQ ID NO: 40, SEQ ID NO: 43, SEQ ID NOs: 48-49, SEQ ID NO: 52, SEQ ID NO: 56, SEQ ID NO: 59, SEQ ID NO: 64, SEQ ID NOs: 68-69, SEQ ID NO: 71 or SEQ ID NO: 337 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule,

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

Another aspect of the present invention provides a method of monitoring for the onset or progression of a neoplasm in an individual, said method comprising measuring the level of expression of one or more *adenoma markers* and/or adenoma markers, as hereinbefore defined, in a biological sample from said individual wherein the level of said *adenoma marker* and/or adenoma marker relative to the normal level of said *adenoma marker* and/or adenoma marker is indicative of the onset of progression of a neoplasm.

In yet another aspect there is provided a method of monitoring for the onset or progression of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more *adenoma markers* and/or adenoma markers, as hereinbefore defined, in a biological sample from said individual wherein the expression profile of said *adenoma markers* and/or adenoma markers relative to normal expression profiles is indicative of the onset or progression of a neoplasm.

Another aspect of the present invention provides a method of classifying an adenoma, said method comprising identifying the expression pattern of one or more *adenoma markers* and/or adenoma markers and/or the expression levels of one or more *adenoma markers* and/or adenoma markers of said adenoma and correlating said adenoma marker expression

results with the morphological and/or phenotypic features of said adenoma.

The present invention should also be understood to extend to the determination of an adenoma's classification status based on the known expression levels and/or expression profiles of the *adenoma markers* and/or adenoma markers expressed by said adenoma, and as previously identified above.

Another aspect of the present invention provides a diagnostic kit for assaying biological samples comprising an agent for detecting one or more *adenoma markers* and/or adenoma markers and reagents useful for facilitating the detection by said agent.

A related aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 1 or SEQ ID NO: 6 or SEQ ID NOs: 8-10 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 1 or SEQ ID NO: 6 or SEQ ID NOs: 8-10 under low stringency conditions at 42°C.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 4 or SEQ ID NO: 5 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 4 or SEQ ID NO: 5 under low stringency conditions at 42°C.

Still yet another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 7 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 7 under low stringency conditions at 42°C.

Yet still another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16 under low stringency conditions at 42°C.

A further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 14 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 14 under low stringency conditions at 42°C.

Another further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23 under low stringency conditions at 42°C.

Yet another further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 20 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 20 under low stringency conditions at 42°C.

Still yet another further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 21 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 21 under low stringency conditions at 42°C.

Yet still another further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 24-26, SEQ ID NO: 31 or SEQ ID NOs: 35-37 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 24-26, SEQ ID NO: 31 or SEQ ID NOs: 35-37 under low stringency conditions at 42°C.

Another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 27, SEQ ID NO: 28 or SEQ ID NO: 29 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 27, SEQ ID NO: 28 or SEQ ID NO: 29 under low stringency conditions at 42°C.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 30 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 30 under low stringency conditions at 42°C.

Still another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 59 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 59 under low stringency conditions at 42°C.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 73-145 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 73-145 under low stringency conditions at 42°C.

A further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336 under low stringency conditions at 42°C.

Yet another aspect of the present invention is directed to an isolated protein selected from the list consisting of:

(i) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 1, SEQ ID NO: 6 or SEQ ID NOs: 8-10 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(ii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 4 or SEQ ID NO: 5 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(iii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in SEQ ID NO: 7 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(iv) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(v) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in SEQ ID NO: 14 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(vi) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(vii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in SEQ ID NO: 20 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(viii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in SEQ ID NO: 21 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(ix) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 24-26, SEQ ID NO: 31 or SEQ ID NOs: 35-37 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(x) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 27 or SEQ ID NO: 28 or SEQ ID NO: 29 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(xi) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence capable of hybridising to any one or more of the nucleotide sequences as set forth in SEQ ID NO: 30 or a derivative, homologue or analogue thereof under low stringency conditions or a derivative, homologue, analogue, chemical equivalent or mimetic of said protein.

(xii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence capable of hybridising to any one or more of the nucleotide sequences as set forth in SEQ ID NO: 59 or a derivative,

homologue or analogue thereof under low stringency conditions or a derivative, homologue, analogue, chemical equivalent or mimetic of said protein.

(xiii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence capable of hybridising to any one or more of the nucleotide sequences as set forth in any one or more of SEQ ID NOs: 73-145 or a derivative, homologue or analogue thereof under low stringency conditions or a derivative, homologue, analogue, chemical equivalent or mimetic of said protein.

(xiv) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence capable of hybridising to any one or more of the nucleotide sequences as set forth in any one or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336 or a derivative, homologue or analogue thereof under low stringency conditions or a derivative, homologue, analogue, chemical equivalent or mimetic of said protein.

The present invention contemplates a method for the treatment and/or prophylaxis of a condition characterised by aberrant, unwanted or otherwise inappropriate cell growth in a subject, said method comprising administering to said subject an effective amount of an agent for a time and under conditions sufficient to modulate *adenoma marker* expression and/or adenoma marker functional activity.

Another aspect of the present invention contemplates the use of an agent as hereinbefore defined in the manufacture of a medicament for the treatment of a condition in a mammal, which condition is characterised by the aberrant, unwanted or otherwise inappropriate cell growth wherein said agent modulates adenoma marker functional activity or *adenoma marker* expression.

In yet another further aspect, the present invention contemplates a pharmaceutical composition comprising a modulatory agent as hereinbefore defined and one or more

pharmaceutically acceptable carriers and/or diluents. Said modulatory agents are referred to as the active ingredients.

Yet another aspect of the present invention relates to modulatory agents, as hereinbefore defined, when used in the method of the present invention.

Still another aspect of the present invention is directed to antibodies to adenoma markers or *adenoma markers* including catalytic antibodies.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a graphical representation of the plot of 8-2d vs 11-10a demonstrating two dimensional cluster analysis;

Figure 2 is a graphical representation of a close up of the graph depicted in Figure 1, highlighting missed tissues using markers 8-2d and 11-10a;

Figure 3 is a graphical representation of tissues missed by all sets of 3 and 4 markers that provide >=69/71 correct classifications;

Figure 4 is a schematic representation of an annotated view of S100P;

Figure 5 is a schematic representation of an annotated view of Defensin α-6;

Figure 6 is a schematic representation of an annotated view of GIF;

Figure 7 is a schematic representation of an annotated view of Reg IV;

Figure 8 is a schematic representation of an annotated view of GW112;

Figure 9 is a schematic representation of an annotated view of Claudin-2;

Figure 10 is a schematic representation of an annotated view of SLC12A2;

Figure 11 is a schematic representation of an annotated view of TGFBI; and

Figure 12 is a schematic representation of an annotated view of Transposon L1.1.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is predicated, in part, on the identification of genetic molecules which have been determined to be expressed at either higher levels or in unique co-expression profiles in individuals who have developed an adenoma than in unaffected individuals. The inventors have still further determined that a proportion of these genetic molecules represent novel genetic molecules. The identification of this population of genetic molecules has now permitted the development of diagnostic methodology based thereon and, further, the identification and rational design of a range of products for use in therapy, prophylaxis, diagnosis and antibody generation.

Accordingly, one aspect of the present invention provides a method for determining the onset or a predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 1-2, SEQ ID NOs: 4-6, SEQ ID NOs: 8-32, SEQ ID NOs: 35-37 or SEQ ID NO: 59 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

In another aspect of the present invention there is provided a method for determining the onset or a predisposition to the onset of a neoplasm in an individual, said method

comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NO: 38, SEQ ID NOs: 40-43, SEQ ID NOs: 45-49, SEQ ID NOs: 58-60, SEQ ID NO: 62, SEQ ID NOs: 64-66, SEQ ID NOs: 68-72 or SEQ ID NOs: 337-338 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

In yet another aspect of the present invention there is provided a method for determining the onset or a predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 73-219 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid

molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

In still another aspect of the present invention there is provided a method for determining the onset or a predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 220-336 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

Reference to "neoplasm" should be understood as a reference to a lesion, tumour or other encapsulated or unencapsulated mass or other form of growth which comprises neoplastic cells. A "neoplastic cell" should be understood as a reference to a cell exhibiting abnormal growth. The term "growth" should be understood in its broadest sense and includes reference to proliferation. In this regard, an example of abnormal cell growth is the uncontrolled proliferation of a cell. The neoplastic cell may be a benign cell or a malignant cell. In a preferred embodiment, the subject neoplasm is an adenoma. Without limiting the present invention to any one theory or mode of action, an adenoma is generally a benign tumour of epithelial origin which is either derived from glandular tissue or exhibits clearly defined glandular structures. It can comprise a malignant cell population within the adenoma, such as occurs in the shift of a benign adenoma to a malignant

adenocarcinoma. Some adenomas exhibit recognisable tissue elements, such a fibrous tissue, while others produce active compounds giving rise to clinical syndrome. Preferably, said adenoma is a gastrointestinal tract adenoma and even more preferably a colorectal adenoma such as a tubular adenoma, tubulovillous adenoma or villous adenoma. Still more preferably, said adenoma is a tubular adenoma, tubulovillous adenoma or villous adenoma greater than 10mm in diameter.

As detailed hereinbefore, it has been determined that modulation in the level of expression or pattern of expression of the nucleic acid molecules detailed above correlates with the development of, or a predisposition to the development of an adenoma, in particular a colorectal adenoma. For ease of reference, these nucleic acid molecules are sometimes herein collectively referred to as "*adenoma markers*". The expression products of the adenoma marker nucleic acid molecules are herein collectively referred to in non-italicised text as "adenoma markers".

Reference to "expression" should be understood as a reference to the transcription and/or translation of a nucleic acid molecule. In this regard, the present invention is exemplified with respect to screening for *adenoma markers* taking the form of mRNA transcripts. Without limiting the present invention in any way, the up-regulation of gene transcription leading to increased mRNA synthesis will also correlate with translation of these mRNA transcripts to produce an expression product. Accordingly, the present invention also extends to adenoma diagnostic methodology which is directed to screening for elevated levels or patterns of expression of the *adenoma marker* expression products as an indicator of the development of, or predisposition to the development of, an adenoma. Although the preferred method is to screen for mRNA transcripts and/or the corresponding expression product, it should be understood that the present invention is not limited in this regard and extends to screening for any other form of *adenoma marker* or its protein expression product such as, for example, a primary RNA transcript. It is well within the skills of the person of skill in the art to determine the most appropriate screening target for any given situation.

Reference to "nucleic acid molecule" should be understood as a reference to both deoxyribonucleic acid molecules and ribonucleic acid molecules. Without limiting the present invention to any one theory or mode of action, the nucleotide sequences disclosed herein are cDNA sequences which correspond to partial or whole mRNA gene transcripts, the concentrations of any one or more of which are elevated greater than 2-fold in individuals exhibiting gastrointestinal tract adenoma development, as compared to unaffected individuals. The present invention therefore extends to both directly screening for mRNA levels in a biological sample or screening for the complimentary cDNA which has been reverse-transcribed from an mRNA population of interest. It is well within the skill of the person of skill in the art to design methodology directed to screening for either DNA or RNA. As detailed above, the method of the present invention also extends to screening for the protein expression product translated from the subject mRNA.

Reference to "biological sample" should be understood as a reference to any sample of biological material derived from an individual such, but not limited to, mucus, stool, urine, blood, serum, biopsy specimens and fluid which has been introduced into the body of an individual and subsequently removed such as, for example, the saline solution extracted from the lung following lung lavage or the solution retrieved from an enema wash. The biological sample which is tested according to the method of the present invention may be tested directly or may require some form of treatment prior to testing. For example, a biopsy sample may require homogenisation prior to testing. To the extent that the neoplasm of interest is a gastrointestinal tract adenoma, the biological sample is preferably a stool sample or any other biological sample of gastrointestinal origin. Where the sample comprises cellular material, it may be necessary to extract or otherwise expose the nucleic acid material present in the cellular material in order to facilitate interaction of a probe with the test sample.

Without limiting the present invention to any one theory or mode of action, it has been determined that, based on a single marker analysis, the *adenoma markers* detailed herein exhibit an upregulation in levels of expression in individuals with adenoma versus those without. The level of upregulation varied from 2 fold to upwards of 200 fold.

Accordingly, in a preferred embodiment the present invention provides a method for determining the onset or the predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecule comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 2 or 30 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

Without limiting the present invention to any one theory or mode of action, it has been determined that these *adenoma markers* are expressed in excess of 100 fold. Further, it has been determined that the nucleic acid molecule defined by SEQ ID NO: 2 corresponds to the gene Claudin 2 and expresses the protein product detailed in SEQ ID NO: 3.

In another embodiment, the present invention provides a method for determining the onset or the predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NO: 7, SEQ ID NO: 14, SEQ ID NOs: 20-21, SEQ ID NOs: 27-29, SEQ ID NO: 38, SEQ ID NO: 43, SEQ ID NO: 49, SEQ ID NOs: 60-62 or SEQ ID NO: 66 or a functional derivative, variant or homologue of said nucleic

acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

Without limiting the present invention to any one theory or mode of action, it has been determined that these *adenoma markers* are expressed at levels of between 10-100 fold above normal levels. Further, it has been determined that a number of the subject nucleic acid molecules correspond to known genes as follows:

- SEQ ID NO: 43 corresponds to the gene encoding gastric intrinsic factor and expresses the protein product detailed in SEQ ID NO: 44.

- SEQ ID NO: 49 corresponds to defensin α-6 (paneth cell specific) and expresses the protein product detailed in SEQ ID NO: 50.

- SEQ ID NO: 66 corresponds to the gene encoding solute carrier family 12, member 2 and expresses the protein product detailed in SEQ ID NO: 67.

- SEQ ID NO: 38 corresponds to the gene encoding regenerating protein IV and expresses the protein product detailed in SEQ ID NO: 39.

- SEQ ID NO: 60 corresponds to the gene encoding GW112 protein and expresses the protein product detailed in SEQ ID NO: 61.

- SEQ ID NO: 62 corresponds to the gene encoding S100 calcium binding protein P and expresses the protein product detailed in SEQ ID NO: 63.

In yet another embodiment, the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NO: 5, SEQ ID NO: 9, SEQ ID NO: 17, SEQ ID NO: 32, SEQ ID NO: 38, SEQ ID NO: 53, SEQ ID NO: 56, SEQ ID NO: 64 or SEQ ID NO: 68 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

Without limiting the present invention to any one theory or mode of action, it has been determined that these *adenoma markers* are expressed at levels between 4.5-10.5 fold above normal levels. Further, it has been determined that the nucleic acid molecule defined by SEQ ID NO: 56 corresponds to the gene transforming growth factor β and expresses the protein product detailed in SEQ ID NO: 57. It has also been determined that the nucleic acid molecule defined by SEQ ID NO: 32 corresponds to the gene encoding transposon L1.1 and expresses the protein product detained in SEQ ID NOs: 33 and 34.

In still another embodiment, the present invention provides a method for determining the

onset or the predisposition to the onset of a neoplasm in an individual, said method comprising measuring the level of expression of one or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NO: 4, SEQ ID NO: 8, SEQ ID NO: 12, SEQ ID NOs: 18-19, SEQ ID NO: 31, SEQ ID NOs: 35-36, SEQ ID NOs: 40-41, SEQ ID NOs: 45-46 or SEQ ID NOs: 51-52, SEQ ID NOs: 54-55, SEQ ID NO: 59, SEQ ID NO: 65, SEQ ID NO: 72 or SEQ ID NOs: 337-338 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any one or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in a biological sample from said individual wherein an increase in the level of expression of said nucleic acid molecule relative to the normal level of expression of said nucleic acid molecule in an individual is indicative of the onset or predisposition to the onset of a neoplasm.

Without limiting the present invention to any one theory or mode of action, it has been determined that these *adenoma markers* are expressed in excess of 1.5-4 fold.

In another most preferred embodiment, the subject nucleotide sequence is SEQ ID NO: 2, SEQ ID NO: 7, SEQ ID NO: 14, SEQ ID NOs: 20-21, SEQ ID NOs: 27-29, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 38, SEQ ID NO: 43, SEQ ID NO: 49, SEQ ID NO: 53, SEQ ID NO: 56, SEQ ID NO: 60, SEQ ID NO: 62 or SEQ ID NO: 66.

In yet another most preferred embodiment the subject nucleotide sequence is SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 20, SEQ ID NOs: 27-29, SEQ ID NO: 38, SEQ ID NO: 43, SEQ ID NO: 49, SEQ ID NO: 53, SEQ ID NO: 56, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID

NO: 66 or SEQ ID NO: 68.

In still yet another preferred embodiment, the subject nucleotide sequence is SEQ ID NO: 2, SEQ ID NO: 7, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 20, SEQ ID NOs: 27-29, SEQ ID NO: 38, SEQ ID NO: 43, SEQ ID NO: 49, SEQ ID NO: 53, SEQ ID NO: 56, SEQ ID NO: 60, SEQ ID NO: 62 or SEQ ID NO: 66.

Most preferably, the subject nucleotide sequence is SEQ ID NO: 2, SEQ ID NO: 7, SEQ ID NO: 14, SEQ ID NO: 43 or SEQ ID NO: 62.

The method of the present invention is predicated on the correlation of levels of *adenoma markers* and/or adenoma markers in individuals with normal levels of said markers. The "normal level" is either the level of *adenoma marker* or adenoma marker in a corresponding biological sample of an individual who has not developed an adenoma nor is predisposed to the development of an adenoma or is the level in a non-adenomous tissue which is derived from the patient who is the subject of testing. This latter method of analysis is a relative form of analysis in terms of the normal and test levels being determined from non-adenomatous and test tissues, respectively, derived from a single individual. However, the method of the present invention should also be understood to encompass non-relative analyses means such as the analysis of test results relative to a standard result which reflects individual or collective results obtained from healthy individuals, other than the patient in issue. Said "normal level" may be a discrete level or a range of levels. Individuals exhibiting *adenoma marker* and/or adenoma marker levels higher than the normal range are generally regarded as having undergone the onset of adenoma development or may be predisposed to the onset of adenoma development.

It should be understood that the "individual" who is the subject of testing may be any human or non-human mammal. Examples of non-human mammals includes primates, livestock animals (e.g. horses, cattle, sheep, pigs, donkeys), laboratory test animals (e.g. mice, rats, rabbits, guinea pigs), companion animals (e.g. dogs, cats) and captive wild animals (e.g. kangaroos, deer, foxes). Preferably the mammal is a human.

In addition to modulation of the level of expression of any one or more *adenoma markers* relative to the levels of expression which are normally observed, the inventors have also surprisingly determined that irrespective of the actual level of expression of any given adenoma marker, the expression, *per se*, of some adenoma markers in combination with other adenoma markers in an individual is indicative of the development of or a predisposition to the development of an adenoma. For example, there can occur similar expression profiles within tissues of the same type, such expression profiles being differentiable from the expression profiles in other tissue types. It should be understood that the level of expression itself is not the unique identifier and may equate to more, less or equal to that which is expressed by healthy individuals. Of relevance is the occurrence of *any* level of expression in combination with other specified markers. The identification of these profile analyses is consistent with current biological understanding that in some situations it is the co-expression of more than one gene which actually causes the development of a given condition. This is certainly consistent with what is known of the complexity of the genomic network. The identification of the diagnostic profiles disclosed herein provides a highly sophisticated means of accurately diagnosing the existence of or predisposition to the development of an adenoma in an individual.

Accordingly, another aspect of the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 1-2, SEQ ID NOs: 4-6, SEQ ID NOs: 8-32, SEQ ID NOs: 35-37 or SEQ ID NO: 59 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

In another aspect, the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NO: 38, SEQ ID NOs: 40-43, SEQ ID NOs: 45-49, SEQ ID NOs: 51-56, SEQ ID NOs: 58-60, SEQ ID NO: 62, SEQ ID NOs: 64-66, SEQ ID NOs: 68-72 or SEQ ID NOs: 337-338 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

In yet another aspect the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 73-219 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

In still another aspect the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 220-336 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

Reference to "co-expression" should be understood as a reference to the simultaneous expression of the subject *adenoma marker* or its expression product. In this regard the onset of co-expression may be simultaneous or it may be staggered. By "staggered" is meant that changes in the gene expression level of one gene occurs at a different time point (ie. either earlier or later) than the change in expression level of other gene or genes. Accordingly, "co-expression" is defined as the subject genes being simultaneously expressed for at least part of the time frame during which each gene is expressed, even if

the subject expression is not commenced and concluded simultaneously. The method of the present invention is directed to detecting these periods of co-expression. It should be understood that although it is preferred that the co-expressed nucleic acid molecules are detectable in the one biological sample, for example a stool sample, they may only be detectable in two separate but simultaneously harvested tissue samples. For example, one adenoma marker may be detectable in a stool sample and the other in a blood sample. It should be understood that such tissue specific expression, which is nevertheless characterised by a period of simultaneous expression, is an example of co-expression within the meaning defined herein.

In a preferred embodiment, the present invention provides a method for determining the onset or predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any three:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NO: 7, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 24, SEQ ID NO: 65, SEQ ID NO: 19, SEQ ID NO: 1, SEQ ID NO: 53, SEQ ID NO: 72, SEQ ID NO: 11 or SEQ ID NO: 26 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any three of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule,

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

Preferably, the subject nucleotide sequences are co-expressed as a profile of three, which profile is selected from the list of:

(i) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 11;
(ii) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 26;
(iii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 16;
(iv) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 1;
(v) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 24; or
(vi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 16.

Still more preferably, the subject nucleotide sequences are co-expressed as a profile of three, which profile is selected from the list of:

(i) SEQ ID NO: 7 and SEQ ID NO: 56 and SEQ ID NO: 11;
(ii) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 11;
(iii) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 11;
(iv) SEQ ID NO: 7 and SEQ ID NO: 9 and SEQ ID NO: 11; or
(v) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 11.

In another preferred embodiment the present invention provides a method for determining the onset or the predisposition to the onset of a neoplasm in an individual, said method comprising detecting the co-expression of any four:

(i) nucleic acid molecules comprising a nucleotide sequence substantially as set forth in any one of SEQ ID NOs: 4-6, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NOs: 21-22, SEQ ID NOs: 27-29, SEQ ID NOs: 30-31, SEQ ID NO: 36, SEQ ID NOs: 37-38, SEQ ID NO: 40, SEQ ID NO: 43, SEQ ID NOs: 48-49, SEQ ID NO: 52, SEQ ID NO: 56, SEQ ID NO: 59, SEQ ID NO: 64, SEQ ID NOs: 68-69, SEQ ID NO: 71 or SEQ ID NO: 337 or a functional derivative, variant or homologue of said nucleic acid molecule; or

(ii) nucleic acid molecules comprising a nucleotide sequence capable of hybridising any two or more of the sequences of (i) under low stringency conditions at 42°C or a functional derivative, variant or homologue of said nucleic acid molecule,

in one or more biological samples from said individual wherein the co-expression of said nucleic acid molecules is indicative of the onset or predisposition to the onset of a neoplasm.

Still more preferably, the subject nucleotide sequences are co-expressed as a profile of four, which profile is selected from the list of:

(i) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 24 and SEQ ID NO: 65;
(ii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 24 and SEQ ID NO: 19;
(iii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 53 and SEQ ID NO: 1;
(iv) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 16 and SEQ ID NO: 19;
(v) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 16 and SEQ ID NO: 46; or
(vi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 65 and SEQ ID NO: 1.

In another most preferred embodiment, the subject nucleotide sequences are co-expressed as a profile of four, which profile is selected from the list of:

(i) SEQ ID NO: 30 and SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 1;
(ii) SEQ ID NO: 7 and SEQ ID NO: 43 and SEQ ID NO: 14 and SEQ ID NO: 24;
(iii) SEQ ID NO: 7 and SEQ ID NO: 43 and SEQ ID NO: 59 and SEQ ID NO: 1;
(iv) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 49 and SEQ ID NO: 24;
(v) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 49 and SEQ ID NO: 16;
(vi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 49 and SEQ ID NO: 1;
(vii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 21 and SEQ ID NO: 16;
(viii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 21 and SEQ ID NO: 1;
(ix) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NOs: 27-29 and SEQ ID NO: 24;
(x) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NOs: 27-29 and SEQ ID NO: 16;
(xi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NOs: 27-29 and SEQ ID NO: 1;
(xii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 56 and SEQ ID NO: 1;
(xiii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 9 and SEQ ID NO: 24;

(xiv) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 9 and SEQ ID NO: 37;
(xv) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 9 and SEQ ID NO: 16;
(xvi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 9 and SEQ ID NO: 1;
(xvii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 24 and SEQ ID NO: 16;
(xviii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 24 and SEQ ID NO: 46;
(xix) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 24 and SEQ ID NO: 1; or
(xx) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 24 and SEQ ID NO: 337.

In yet another most preferred embodiment, the subject nucleotide sequences are co-expressed as a profile of four, which profile is selected from the list of:

(i) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 5 and SEQ ID NO: 1;
(ii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 65 and SEQ ID NO: 16;
(iii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 65 and SEQ ID NO: 1;
(iv) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 53 and SEQ ID NO: 37;
(v) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 53 and SEQ ID NO: 48;
(vi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 68 and SEQ ID NO: 1;
(vii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 31 and SEQ ID NO: 1;
(viii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 69 and SEQ ID NO: 16;
(ix) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 69 and SEQ ID NO: 1;
(x) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 52 and SEQ ID NO: 1;
(xi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 16 and SEQ ID NO: 337;
(xii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 16 and SEQ ID NO: 71;
(xiii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 36 and SEQ ID NO: 1;
(xiv) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 19 and SEQ ID NO: 1;
(xv) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 40 and SEQ ID NO: 1;
(xvi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 22 and SEQ ID NO: 1;
(xvii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 46 and SEQ ID NO: 1;
(xviii) SEQ ID NO: 7 and SEQ ID NOs: 27-29 and SEQ ID NO: 24 and SEQ ID NO: 4;
(xix) SEQ ID NO: 7 and SEQ ID NOs: 27-29 and SEQ ID NO: 65 and SEQ ID NO: 11; or

(xx) SEQ ID NO: 7 and SEQ ID NO: 38 and SEQ ID NO: 64 and SEQ ID NO: 13.

In still another most preferred embodiment, the subject nucleotide sequences are co-expressed as a profile of four, which profile is selected from the list of:

(i) SEQ ID NO: 7 and SEQ ID NO: 9 and SEQ ID NO: 68 and SEQ ID NO: 11;
(ii) SEQ ID NO: 7 and SEQ ID NO: 24 and SEQ ID NO: 69 and SEQ ID NO: 11;
(iii) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 53 and SEQ ID NO: 11;
(iv) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 68 and SEQ ID NO: 11;
(v) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 69 and SEQ ID NO: 13;
(vi) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 36 and SEQ ID NO: 13;
(vii) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 11 and SEQ ID NO: 337;
(viii) SEQ ID NO: 7 and SEQ ID NO: 53 and SEQ ID NO: 72 and SEQ ID NO: 11;
(ix) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 26 and SEQ ID NO: 46;
(x) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 36 and SEQ ID NO: 11;
(xi) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 46 and SEQ ID NO: 11;
(xii) SEQ ID NO: 7 and SEQ ID NO: 69 and SEQ ID NO: 46 and SEQ ID NO: 11;
(xiii) SEQ ID NO: 43 and SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 24;
(xiv) SEQ ID NO: 43 and SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 16;
(xv) SEQ ID NO: 43 and SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 1;
(xvi) SEQ ID NO: 43 and SEQ ID NO: 7 and SEQ ID NOs: 27-29 and SEQ ID NO: 24;
(xvii) SEQ ID NO: 43 and SEQ ID NO: 7 and SEQ ID NO: 36 and SEQ ID NO: 11;
(xviii) SEQ ID NO: 43 and SEQ ID NO: 7 and SEQ ID NO: 59 and SEQ ID NO: 1;
(xix) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 49 and SEQ ID NO: 24; or
(xx) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 49 and SEQ ID NO: 22.

In yet still another most preferred embodiment, the subject nucleotide sequences are co-expressed as a profile of four, which profile is selected from the list of:

(i) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 49 and SEQ ID NO: 1;
(ii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 56 and SEQ ID NO: 1;

(iii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 56 and SEQ ID NO: 1;
(iv) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 9 and SEQ ID NO: 1;
(v) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 24 and SEQ ID NO: 19;
(vi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 65 and SEQ ID NO: 37;
(vii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 53 and SEQ ID NO: 48;
(viii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 53 and SEQ ID NO: 1;
(ix) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 72 and SEQ ID NO: 1;
(x) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 69 and SEQ ID NO: 16;
(xi) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 16 and SEQ ID NO: 19;
(xii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 19 and SEQ ID NO: 1;
(xiii) SEQ ID NO: 7 and SEQ ID NO: 14 and SEQ ID NO: 1 and SEQ ID NO: 71;
(xiv) SEQ ID NO: 7 and SEQ ID NO: 49 and SEQ ID NO: 64 and SEQ ID NO: 11;
(xv) SEQ ID NO: 7 and SEQ ID NO: 38 and SEQ ID NO: 56 and SEQ ID NO: 13;
(xvi) SEQ ID NO: 7 and SEQ ID NO: 38 and SEQ ID NO: 56 and SEQ ID NO: 13;
(xvii) SEQ ID NO: 7 and SEQ ID NO: 56 and SEQ ID NO: 64 and SEQ ID NO: 11;
(xviii) SEQ ID NO: 7 and SEQ ID NO: 56 and SEQ ID NO: 53 and SEQ ID NO: 6;
(xix) SEQ ID NO: 7 and SEQ ID NO: 9 and SEQ ID NO: 64 and SEQ ID NO: 16; or
(xx) SEQ ID NO: 7 and SEQ ID NO: 9 and SEQ ID NO: 64 and SEQ ID NO: 13.

In still yet another most preferred embodiment, the subject nucleotide sequences are co-expressed as a profile of four, which profile is selected from the list of:

(i) SEQ ID NO: 7 and SEQ ID NO: 9 and SEQ ID NO: 68 and SEQ ID NO: 11;
(ii) SEQ ID NO: 7 and SEQ ID NO: 24 and SEQ ID NO: 72 and SEQ ID NO: 13;
(iii) SEQ ID NO: 7 and SEQ ID NO: 24 and SEQ ID NO: 72 and SEQ ID NO: 46;
(iv) SEQ ID NO: 7 and SEQ ID NO: 24 and SEQ ID NO: 72 and SEQ ID NO: 71;
(v) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 72 and SEQ ID NO: 16;
(vi) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 68 and SEQ ID NO: 11;
(vii) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 69 and SEQ ID NO: 11;
(viii) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 19 and SEQ ID NO: 11;
(ix) SEQ ID NO: 7 and SEQ ID NO: 64 and SEQ ID NO: 13 SEQ ID NO: 11;

(x) SEQ ID NO: 7 and SEQ ID NO: 53 and SEQ ID NO: 72 and SEQ ID NO: 11;
(xi) SEQ ID NO: 7 and SEQ ID NO: 53 and SEQ ID NO: 15 and SEQ ID NO: 11;
(xii) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 68 and SEQ ID NO: 11;
(xiii) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 69 and SEQ ID NO: 11;
(xiv) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 36 and SEQ ID NO: 11;
(xv) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 19 and SEQ ID NO: 11;
(xvi) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 46 and SEQ ID NO: 11;
(xvii) SEQ ID NO: 7 and SEQ ID NO: 72 and SEQ ID NO: 46 and SEQ ID NO: 1;
(xviii) SEQ ID NO: 7 and SEQ ID NO: 68 and SEQ ID NO: 16 and SEQ ID NO: 36; or
(xix) SEQ ID NO: 7 and SEQ ID NO: 68 and SEQ ID NO: 36 and SEQ ID NO: 11.

In yet still another preferred embodiment, the subject nucleotide sequences are co-expressed as a profile of five.

In a further preferred embodiment, the subject nucleotide sequences are co-expressed as a profile of six.

In still yet a further preferred embodiment, the subject nucleotide sequences are co-expressed as a profile of more than 6.

In accordance with these preferred embodiments, the subject neoplasm is preferably an adenoma and even more preferably a colorectal adenoma.

As detailed hereinbefore, it should be understood that although the present invention is exemplified with respect to the detection of nucleic acid molecules, it also encompasses methods of detection based on screening for the expression product of the subject *adenoma markers* or derivatives thereof. The present invention should also be understood to mean methods of screening based on identifying either protein product and nucleic acid material in one or more biological samples. However, it should be understood that some of the *adenoma markers* may correlate to genes or gene fragments which do not encode a protein expression product. Accordingly, to the extent that this occurs it would not be possible to

screen for an expression product and the subject marker must be assessed on the basis of nucleic acid expression profiles.

Without limiting the present invention in any way, the following expression products are exemplified herein:

(i) SEQ ID NO: 2 corresponds to the claudin 2 gene and encodes the expression product detailed in SEQ ID NO: 3;

(ii) SEQ ID NO: 32 corresponds to the transposon L1.1 gene and encodes the expression product detailed in SEQ ID NOs: 33 and 34;

(iii) SEQ ID NO: 38 corresponds to the regenerating protein IV gene and encodes the expression product detailed in SEQ ID NO: 39;

(iv) SEQ ID NO: 43 corresponds to the gastric intrinsic factor gene and encodes the expression product detailed in SEQ ID NO: 44;

(v) SEQ ID NO: 49 corresponds to the defensin α6 gene (paneth cell specific) and encodes the expression product detailed in SEQ ID NO: 50;

(vi) SEQ ID NO: 56 corresponds to the TGF β1 gene and encodes the expression product detailed in SEQ ID NO: 57;

(vii) SEQ ID NO: 60 corresponds to the GW112 gene and encodes the expression product detailed in SEQ ID NO: 61

(viii) SEQ ID NO: 62 corresponds to the S100P gene and encodes the expression product detailed in SEQ ID NO: 63;

(ix) SEQ ID NO: 66 corresponds to the SLC12A1 gene and encodes the expression product detailed in SEQ ID NO: 67.

Further details in relation to each of these genes and proteins are provided in Example 6.

"Derivatives" should be understood to have the same meaning as hereinafter provided. Reference to a "functional derivative" should be understood as a reference to a derivative which, in accordance with the teachings provided herein, is indicative of the development of a neoplasm, in particular adenoma. In particular, however, the subject derivative may

be a partially degraded or denatured molecule. For example, the mRNA which is screened for in stool samples in accordance with the exemplification provided herein is likely to lack the polyA tail which generally characterises a mRNA transcript. In another example, proteinaceous adenoma markers may be fragmented, denatured (for example due to breakdown of disulphide bonds) or may be otherwise degraded. This is likely to be the case, for example, where the biological sample which is the subject of screening comprises proteinases, such as are sometimes found in urine.

Reference herein to *adenoma markers* or adenoma markers (either collectively or in terms of specific SEQ ID NOs) should be read as including reference to all forms of these molecules and to functional derivatives, variants or homologues thereof, in the context of the diagnostic aspects of the present invention. Accordingly, reference to *adenoma markers* should be understood to include reference to isoforms which arise from alternative splicing of the adenoma marker mRNA or mutants or polymorphic variants of the *adenoma markers*. In this regard, for example, it is particularly significant to note that the markers exemplified herein have been derived from individual tissue. However, some genes are known to exhibit allelic variation between individuals. Accordingly, the present invention should be understood to extend to such variants which, in terms of the concept of the present diagnostic applications, achieve the same outcome despite the fact that minor genetic variants between the actual nucleic acid sequences may exist between individuals. Accordingly, the present invention should be understood to extend to all mRNA, cDNA and peptide isoforms which arise from alternative splicing or any other mutation or polymorphic variation.

Reference to the "onset" of a neoplasm, preferably adenoma development, should be understood as a reference to one or more cells of that individual exhibiting abnormal growth characteristic. In this regard, the adenoma may be well developed in that a mass of proliferating cells has developed. Alternatively, the adenoma may be at a very early stage in that only relatively few abnormal cell divisions have occurred at the time of diagnosis. The present invention also extends to the assessment of an individual's predisposition to the development of a neoplasm, such as an adenoma. Without limiting the present

invention in any way, increased levels of or expression profiles of *adenoma markers* or adenoma markers in an individual who has not undergone the onset of adenoma development may be indicative of that individual's predisposition to developing an adenoma, such as the imminent development of an adenoma.

Although the preferred method is to detect the expression of *adenoma markers* and/or adenoma markers for the purpose of diagnosing adenoma development or predisposition thereto, the detection of a decrease in the levels of or down-regulation of expression profiles said markers may be desired under certain circumstances, for example, to monitor the effectiveness of therapeutic or prophylactic treatment directed to modulating a neoplastic condition, such as adenoma development. For example, where elevated levels of *adenoma markers* and/or adenoma markers indicated that an individual had developed a condition characterised by adenoma development, screening for a decrease in the levels of these markers subsequently to the onset of a therapeutic regime may be utilised to indicate reversal or other form of improvement of the subject individual's condition.

The method of the present invention is useful as a one off test or as an on-going monitor of those individuals thought to be at risk of adenoma development or as a monitor of the effectiveness of therapeutic or prophylactic treatment regimes directed to inhibiting or otherwise slowing adenoma development. In these situations, mapping the modulation of *adenoma marker* and/or adenoma marker levels or expression profiles in any one or more classes of biological samples is a valuable indicator of the status of an individual or the effectiveness of a therapeutic or prophylactic regime which is currently in use. Accordingly, the method of the present invention should be understood to extend to monitoring for increases or decreases in marker levels or expression profiles in an individual relative to their normal level (as hereinbefore defined) or relative to one or more earlier marker levels or expression profiles determined from a biological sample of said individual.

Accordingly, another aspect of the present invention provides a method of monitoring for the onset or progression of a neoplasm in an individual, said method comprising measuring

the level of expression of one or more *adenoma markers* and/or adenoma markers, as hereinbefore defined, in a biological sample from said individual wherein the level of said *adenoma marker* and/or adenoma marker relative to the normal level of said *adenoma marker* and/or adenoma marker is indicative of the onset of progression or a neoplasm.

In yet another aspect there is provided a method of monitoring for the onset or progression of a neoplasm in an individual, said method comprising detecting the co-expression of any two or more *adenoma markers* and/or adenoma markers, as hereinbefore defined, in a biological sample from said individual wherein the expression profile of said *adenoma markers* and/or adenoma markers relative to normal expression profiles is indicative of the onset or progression of a neoplasm.

Preferably said neoplasm is an adenoma. Even more preferably said adenoma is a gastrointestinal tract adenoma. Most preferably said gastrointestinal tract adenoma is a colorectal adenoma.

In still another aspect, the present invention extends to the classification of adenomas obtained by biopsy based on the expression profile and/or expression levels of one or more of the *adenoma markers* and/or adenoma markers defined herein.

Without limiting the present invention to any one theory or mode of action, adenomas may develop through progressive stages of size, appearance and dysplasia (cellular disorganization), into colorectal cancer (Young, G.P., Rozen, P. and Levin, B. Chapter 3: Ed. Rozen, P., Young, G.P., Levin, P., Spann, S.J. Martin Dunitz 2002). The process is by a series of steps, each of which constitutes a change in the biology that is driven by accumulation of genetic mutations. Unless inherited, the mutations occur by chance and in random order. Once acquired the process is not inevitable, as more chance mutations must occur.

Progression of adenoma development is characterised by any one or more of the following:

a) An increase in size: 1cm or over is definitely at-risk and some consider >5mm as significant.

b) An increase in the villous component.
- The usual histology is "tubular". Adenomas can develop a villous appearance in some areas. If <25%, the adenoma remains a tubular adenoma (TA).
- 25-50% is a tubulovillous adenoma (TVA).
- >50% change is a villous adenoma (VA). Regardless of size, villous change is important and risk increases as the degree of villous change increases.

c) An increase in the degree of dysplasia. All adenomas show dysplasia by definition, as this is the histologic hallmark of neoplasia. Adenomas are usually classified as a low (LGD) or high (HGD) grade dysplasia. High-grade means increased risk.

d) Multiplicity. While not strictly a marker of progression for the individual adenoma, multiplicity does denote increased risk of progression.

The risk of a polyp developing into cancer is summarized in the following table:

Type of polyp	**Risk increase**
Adenoma: Multiple (*i.e.* >2 of any size)	8-fold
Adenoma: >9mm, or villous change, or HGD	4-fold
Adenoma: <10mm, tubular and single	No identified increased risk
Hyperplastic	No identified increased risk

Many adenomas never progress, but remain single, tubular and less than 1cm in size. Overall, it is estimated that about 5-10% of adenomas will progress to cancer. The process is generally slow and takes about 10 years. The adenoma "dwell-time" takes 5-10 years and a cancer may take 5 years before death ensues. Most adenomas are asymptomatic, as

are cancers in the early phases.

Accordingly, the present invention now provides a means of classifying adenomas other than by the currently accepted techniques which are based on gross histological morphology. The currently used methods lack precision in relation to disease classification and prognosis. The identification of *adenoma markers* and adenoma markers together with identification of their expression uplift levels and expression profile can now be correlated to disease stage and/or cancer invasiveness. This provides a means of more effectively classifying, staging and predicting disease progression.

Accordingly, another aspect of the present invention provides a method of classifying an adenoma, said method comprising identifying the expression pattern of one or more *adenoma markers* and/or adenoma markers and/or the expression levels of one or more *adenoma markers* and/or adenoma markers of said adenoma and correlating said adenoma marker expression results with the morphological and/or phenotypic features of said adenoma.

The present invention should also be understood to extend to the determination of an adenoma's classification status based on the known expression levels and/or expression profiles of the *adenoma markers* and/or adenoma markers expressed by said adenoma, and as previously identified above.

Means of screening for *adenoma markers* or adenoma markers in a biological sample can be achieved by any suitable method, which would be well known to the person of skill in the art, such as but not limited to:

(i) *In vivo* detection of adenomas.

Molecular Imaging may be used following administration of imaging probes or reagents capable of disclosing altered expression of the markers in the intestinal tissues.

Molecular imaging (Moore, A., Basilion, J., Chiocca, e., and Weissleder, R., Measuring Transferrin Receptor Gene Expression by NMR Imaging. BBA, 1402:239-249, 1988; Weissleder, R., Moore, A., Ph.D., Mahmood-Bhorade, U., Benveniste, H., Chiocca, E.A., Basilion, J.P. High resolution *in vivo* imaging of transgene expression, Nature Medicine, 6:351-355, 2000) is the *in vivo* imaging of molecular expression that correlates with the macro-features currently visualized using "classical" diagnostic imaging techniques such as X-Ray, computed tomography (CT), MRI, Positron Emission Tomography (PET) or endoscopy. Historically, detection of malignant tumor cells in a background of normal or hyperplastic benign tissue is often based on differences in physical properties between tissues, which are frequently minimal, resulting in low contrast resolution. Application of expression profiling will define the differences in "molecular properties" between cancer and normal tissues that arise as a result of malignant transformation. Definition of "molecular signatures" for adenomas will enable development of more sensitive and informative imaging methods that exploit these genetic differences. Additionally, identification of imaging marker genes, expression profiles or gene products for adenomas that can be correlated to disease stage and/or cancer invasiveness may eventually lead to non-invasive staging and prognosis of disease.

(ii) Detection of up-regulation of mRNA expression in the cells by Fluorescent *In Situ* Hybridization (FISH), or in extracts from the cells by technologies such as Quantitative Reverse Transcriptase Polymerase Chain Reaction (QRTPCR) or Flow cytometric qualification of competitive RT-PCR products (Wedemeyer, N., Potter, T., Wetzlich, S. and Gohde, W. Flow Cytometric Quantification of Competitive Reverse Transcriptase-PCR products, Clinical Chemistry 48:9 1398-1405, 2002).

(iii) Assessment of expression profiles of mRNA from cellular extracts, for example by array technologies (Alon, A., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mach, D. and Levine, A.J. Broad patterns of gene expression revealed by clustering

analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. USA: 96, 6745-6750, June 1999).

(iv) Measurement of altered marker protein levels in cell extracts, for example by immunoassay.

(v) The use of aptamers in screening for nucleic acid molecules or expression products

(vi) Determining altered expression of protein markers on the cell surface, for example by immunohistochemistry.

(vii) Determining altered protein expression based on any suitable functional test, enzymatic test or immunological test in addition to those detailed in points (iv) and (vi) above.

A person of ordinary skill in the art could determine, as a matter of routine procedure, the appropriateness of applying a given method to a particular type of biological sample.

Another aspect of the present invention provides a diagnostic kit for assaying biological samples comprising an agent for detecting one or more *adenoma markers* and/or adenoma markers and reagents useful for facilitating the detection by the agent in the first compartment. Further means may also be included, for example, to receive a biological sample. The agent may be any suitable detecting molecule.

As detailed hereinbefore, the inventors have determined that a proportion of the *adenoma markers* represent novel genetic molecules. The identification of this population of genes has now permitted the rational design of a range of products and methods for use in diagnosis, therapy, prophylaxis and antibody generation.

Accordingly, a related aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence

substantially as set forth in any one or more of SEQ ID NO: 1 or SEQ ID NO: 6 or SEQ ID NOs: 8-10 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 1 or SEQ ID NO: 6 or SEQ ID NOs: 8-10 under low stringency conditions at 42°C.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 4 or SEQ ID NO: 5 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 4 or SEQ ID NO: 5 under low stringency conditions at 42°C.

Still yet another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 7 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 7 under low stringency conditions at 42°C.

Yet still another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16 under low stringency conditions at 42°C.

A further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 14 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 14 under low stringency conditions at 42°C.

Another further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence

substantially as set forth in any one or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23 under low stringency conditions at 42°C.

Yet another further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 20 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 20 under low stringency conditions at 42°C.

Still yet another further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 21 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 21 under low stringency conditions at 42°C.

Yet still another further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 24-26, SEQ ID NO: 31 or SEQ ID NOs: 35-37 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 24-26, SEQ ID NO: 31 or SEQ ID NOs: 35-37 under low stringency conditions at 42°C.

Another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 27, SEQ ID NO: 28 or SEQ ID NO: 29 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 27, SEQ ID NO: 28 or SEQ ID NO: 29 under low stringency conditions at 42°C.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence

substantially as set forth in any one or more of SEQ ID NO: 30 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 30 under low stringency conditions at 42°C.

Still another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 59 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NO: 59 under low stringency conditions at 42°C.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 73-145 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 73-145 under low stringency conditions at 42°C.

A further aspect of the present invention contemplates an isolated nucleic acid molecule or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336 or derivative or homologue thereof, or capable of hybridising to any one or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336 under low stringency conditions at 42°C.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01M to at least about 0.15M salt for

hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions. In general, washing is carried out at $T_m = 69.3 + 0.41$ (G + C) % [19] = -12°C. However, the T_m of a duplex DNA decreases by 1°C with every increase of 1% in the number of mismatched based pairs (Bonner *et al* (1973) *J. Mol. Biol.* **81**:123).

Without limiting the present invention to any one theory or mode of action, the nucleic acid molecules according to these aspects of the present invention are cDNA sequences which correspond to partial mRNA gene transcripts, the concentrations of which are elevated more than 2-fold in the tissues of individuals exhibiting gastrointestinal tract adenoma development, as compared to unaffected individuals or tissues. These mRNA transcript sequences do not correlate with any known gene sequences and therefore reflect the identification of a novel population of genes which, *inter alia*, are useful as diagnostic markers of adenoma development. More particularly, these genes are thought to form a novel group of genes which are indicative of the onset of or a predisposition to the onset of adenoma development when expressed at either higher levels or in unique co-expression profiles in individuals who have developed an adenoma than in unaffected individuals. The nucleic acid molecules according to this aspect of the present invention are herein collectively referred to as "*novel adenoma markers*". The expression product of the adenoma marker nucleic acid molecules are herein referred to non-italicised text as "novel adenoma markers". It should be understood that these molecules form a subgroup of the "*adenoma markers*" and "adenoma markers" as defined in relation to the diagnostic aspects of the present invention. For ease of reference, the two groups of markers are distinguishable by the presence or absence of the prefix "novel".

More particularly, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NO: 1 or SEQ ID NO: 6 or SEQ ID NOs: 8-10.

In another embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NO: 4 or SEQ ID NO: 5.

In yet another embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 7.

In yet another embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16.

In still another embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 14.

In yet still another embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23.

In still yet another embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 20.

In a further embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 21.

In a still further embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 24-26, SEQ ID NO: 31 or SEQ ID NOs: 35-37.

In a yet further embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NO: 27, SEQ ID NO: 28 or SEQ ID NO: 29.

In a yet still further embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NO: 30.

In yet another embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NO: 59.

In still another embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 73-145.

In a further embodiment, the present invention contemplates a nucleic acid molecule or a derivative, homologue or analogue thereof comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336.

The *novel adenoma marker* nucleic acid molecules of the present invention are preferably cDNA sequences of deoxyribonucleic acids or mRNA sequences of ribonucleic acids. However, the novel nucleic acid molecules of the present invention should be understood to extend to any form of deoxyribonucleic acid or ribonucleic acid molecule such as

genomic sequences (which will comprise exons and introns and may also comprise promoter or other regulatory regions) or primary RNA transcript sequences.

Accordingly, another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to any one or more of SEQ ID NO: 1 or SEQ ID NO: 6 or SEQ ID NOs: 8-16 under low stringency conditions at 42°C.

Yet another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to any one or more of SEQ ID NO: 4 or SEQ ID NO: 5 under low stringency conditions at 42°C.

Still yet another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to SEQ ID NO: 7 under low stringency conditions at 42°C.

Yet still another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to any one or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16 under low stringency conditions at 42°C.

A further aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to SEQ ID NO: 14 under low stringency conditions at 42°C.

Another further aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to any one or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23 under low stringency conditions at 42°C.

Yet another further aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to SEQ ID NO: 20 under low stringency conditions at 42°C.

Still yet another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to SEQ ID NO: 21 under low stringency conditions at 42°C.

Yet still another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to any one or more of SEQ ID NOs: 24-26, SEQ ID NO: 31 or SEQ ID NOs: 35-37 under low stringency conditions at 42°C.

Another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to any one or more of SEQ ID NOs: 27-29 under low stringency conditions at 42°C.

Yet another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to SEQ ID NO: 30 under low stringency conditions at 42°C.

Still another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to SEQ ID NO: 59 under low stringency conditions at 42°C.

Yet another aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to any one or more of SEQ ID NOs: 73-145 under low stringency conditions at 42°C.

A further aspect of the present invention contemplates a genomic nucleic acid molecule or derivative, homologue or analogue thereof capable of hybridising to any one or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336 under low stringency conditions at 42°C.

Reference herein to "*novel adenoma markers*" and "novel adenoma markers" should be understood as a reference to all forms of these molecules and to derivatives, homologues, analogues, chemical equivalents and mimetics thereof including, for example, any peptide or cDNA isoforms which arise from alternative splicing of *novel adenoma marker* mRNA or mutants or polymorphic variants of *novel adenoma markers* or novel adenoma markers.

The molecules disclosed herein have been isolated from the human. However, it should be understood that the protein and/or nucleic acid molecules may also be isolated from any other animal or non-animal source. For example, other animal and non-animal sources include, but are not limited to, primates, livestock animals (e.g. sheep, pigs, cows, goats, horses, donkeys), laboratory test animals (e.g. mice, hampsters, rabbits, rats, guinea pigs), domestic companion animals (e.g. dogs and cats), birds (e.g. chicken, geese, ducks and other poultry birds, game birds, emus, ostriches), captive wild or tamed animals (e.g. foxes, kangaroos, dingoes), reptiles or fish.

The nucleic acid molecule of the present invention is preferably in isolated form or ligated to a vector, such as an expression vector. By "isolated" is meant a nucleic acid molecule having undergone at least one purification step and this is conveniently defined, for example, by a composition comprising at least about 10% subject nucleic acid molecule, preferably at least about 20%, more preferably at least about 30%, still more preferably at least about 40-50%, even still more preferably at least about 60-70%, yet even still more preferably 80-90% or greater of subject nucleic acid molecule relative to other components as determined by molecular weight, encoding activity, nucleotide sequence, base composition or other convenient means. The nucleic acid molecule of the present invention may also be considered, in a preferred embodiment, to be biologically pure.

In a particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NO: 1 or SEQ ID NO: 6 or SEQ ID NOs: 8-10 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to anyone or more of SEQ ID NO: 1 or SEQ ID NO: 6 or SEQ ID NOs: 8-10.

In another particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NO: 4 or SEQ ID NO: 5 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to anyone or more of SEQ ID NO: 4 or SEQ ID NO: 5.

In yet another particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 7 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to SEQ ID NO: 7.

In still another particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to anyone or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16.

In yet still another particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 14 or is a derivative, homologue or

analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to SEQ ID NO: 14.

In still yet another particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to anyone or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23.

In a further particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 20 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to SEQ ID NO: 20.

In a still further particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 21 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to SEQ ID NO: 21.

In a yet further particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 24-26 or SEQ ID NO: 31 or SEQ ID NOs: 35-37 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to anyone or more of SEQ ID NOs: 24-26 or SEQ ID NO: 31 or SEQ ID NOs: 35-37.

In a still yet further particularly preferred embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of

nucleotides substantially as set forth in any one or more of SEQ ID NOs: 27-29 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to anyone or more of SEQ ID NOs: 27-29.

In another embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 30 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to SEQ ID NO: 30.

In yet another embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in SEQ ID NO: 59 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to SEQ ID NO: 59.

In still another embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 73-145 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to anyone or more of SEQ ID NOs: 73-145.

In a further embodiment, the nucleotide sequence corresponding to a *novel adenoma marker* is a cDNA sequence comprising a sequence of nucleotides substantially as set forth in any one or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336 or is a derivative, homologue or analogue thereof including a cDNA sequence comprising a sequence of nucleotides having similarity to anyone or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336.

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or

conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. The percentage similarity may be greater than 50% such as at least 70% or at least 80% or at least 90% or at least 95% or higher.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences may be aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions can then be compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e. % identity = # of identical positions/total # of overlapping positions x 100). Preferably, the two sequences are the same length. The determination of percent identity or homology between two sequences can be accomplished using a mathematical algorithm. A suitable, mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990), modified as in Karlin and Altschul (1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, *et al.* (1990) *J. Mol. Biol. 215:*403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. Another example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software

package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The nucleic acid molecule may be ligated to an expression vector capable of expression in a prokaryotic cell (e.g. *E. coli*) or a eukaryotic cell (e.g. yeast cells, fungal cells, insect cells, mammalian cells or plant cells). The nucleic acid molecule may be ligated or fused or otherwise associated with a nucleic acid molecule encoding another entity such as, for example, a signal peptide, epitope tag, fluorescent tag, dimerisation motif, inhibitory motif, activation motif or regulatory motif.

The present invention extends to the expression product of the nucleic acid molecules hereinbefore defined.

Accordingly, yet another aspect of the present invention is directed to an isolated protein selected from the list consisting of:

(i) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 1, SEQ ID NO: 6 or SEQ ID NOs: 8-10 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(ii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 4 or SEQ ID NO: 5 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(iii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in SEQ ID NO:

7 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(iv) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 11-13 or SEQ ID NOs: 15-16 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(v) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in SEQ ID NO: 14 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(vi) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 17-19 or SEQ ID NOs: 22-23 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(vii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in SEQ ID NO: 20 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(viii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in SEQ ID NO: 21 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(ix) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NOs: 24-26, SEQ ID NO: 31 or SEQ ID NOs: 35-37 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(x) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence substantially as set forth in any one or more of SEQ ID NO: 27 or SEQ ID NO: 28 or SEQ ID NO: 29 or a derivative, homologue, analogue, chemical equivalent or mimetic or said protein.

(xi) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence capable of hybridising to any one or more of the nucleotide sequences as set forth in SEQ ID NO: 30 or a derivative, homologue or analogue thereof under low stringency conditions or a derivative, homologue, analogue, chemical equivalent or mimetic of said protein.

(xii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence capable of hybridising to any one or more of the nucleotide sequences as set forth in SEQ ID NO: 59 or a derivative, homologue or analogue thereof under low stringency conditions or a derivative, homologue, analogue, chemical equivalent or mimetic of said protein.

(xiii) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence capable of hybridising to any one or more of the nucleotide sequences as set forth in any one or more of SEQ ID NOs: 73-145 or a derivative, homologue or analogue thereof under low stringency conditions or a derivative, homologue, analogue, chemical equivalent or mimetic of said protein.

(xiv) a protein encoded by a nucleotide sequence or derivative, homologue or analogue thereof comprising a nucleotide sequence capable of hybridising to any one or more of the nucleotide sequences as set forth in any one or more of SEQ ID NOs: 146-219 or SEQ ID NO: 336 or a derivative, homologue or analogue thereof under low stringency conditions or a derivative, homologue, analogue, chemical equivalent or mimetic of said protein.

The term "protein" should be understood to encompass peptides, polypeptides and proteins. The protein may be glycosylated or unglycosylated and/or may contain a range

of other molecules fused, linked, bound or otherwise associated to the protein such as amino acids, lipids, carbohydrates or other peptides, polypeptides or proteins. Reference herein to a "protein" includes a protein comprising a sequence of amino acids as well as a protein associated with other molecules such as amino acids, lipids, carbohydrates or other peptides, polypeptides or proteins.

The protein of the present invention is preferably in isolated form. By "isolated" is meant a protein having undergone at least one purification step and this is conveniently defined, for example, by a composition comprising at least about 10% subject protein, preferably at least about 20%, more preferably at least about 30%, still more preferably at least about 40-50%, even still more preferably at least about 60-70%, yet even still more preferably 80-90% or greater of subject protein relative to other components as determined by molecular weight, amino acid sequence or other convenient means. The protein of the present invention may also be considered, in a preferred embodiment, to be biologically pure.

The adenoma marker proteins of the present invention may be in multimeric form meaning that two or more molecules are associated together. Where the same protein molecules are associated together, the complex is a homomultimer. An example of a homomultimer is a homodimer. Where at least one adenoma marker protein is associated with at least one non-adenoma marker protein, then the complex is a heteromultimer such as a heterodimer.

The ability to produce recombinant proteins permits the large scale production of adenoma markers for commercial use. The adenoma markers may need to be produced as part of a large peptide, adenomaeptide or protein which may be used as is or may first need to be processed in order to remove the extraneous proteinaceous sequences. Such processing includes digestion with proteases, peptidases and amidases or a range of chemical, electrochemical, sonic or mechanical disruption techniques.

Notwithstanding that the present invention encompasses recombinant proteins, chemical synthetic techniques are also preferred in the synthesis of the subject proteins.

Adenoma marker proteins according to the present invention are conveniently synthesised based on molecules isolated from the human. Isolation of the human molecules may be accomplished by any suitable means such as by chromotographic separation, for example using CM-cellulose ion exchange chromotography followed by Sephadex (e.g. G-50 column) filtration. Many other techniques are available including HPLC, PAGE amongst others.

The subject proteins may be synthesised by solid phase synthesis using F-moc chemistry as described by Carpino *et al.* (1991). Proteins and fragments thereof may also be synthesised by alternative chemistries including, but not limited to, t-Boc chemistry as described in Stewart *et al.* (1985) or by classical methods of liquid phase peptide synthesis.

Derivatives of the nucleic acid and protein molecules defined herein include fragments, parts, portions, mutants, variants and mimetics from natural, synthetic or recombinant sources including fusion proteins. Parts or fragments include, for example, active regions of the adenoma markers. Derivatives may be derived from insertion, deletion or substitution of amino acids. Amino acid insertional derivatives include amino and/or carboxylic terminal fusions as well as intrasequence insertions of single or multiple amino acids. Insertional amino acid sequence variants are those in which one or more amino acid residues are introduced into a predetermined site in the protein although random insertion is also possible with suitable screening of the resulting product. Deletional variants are characterized by the removal of one or more amino acids from the sequence. Substitutional amino acid variants are those in which at least one residue in the sequence has been removed and a different residue inserted in its place. An example of substitutional amino acid variants are conservative amino acid substitutions. Conservative amino acid substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid and glutamic acid; asparagine and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine. Additions to amino acid sequences include fusions with other peptides, polypeptides or proteins.

Homologues should be understood as a reference to nucleic acid or protein molecules isolated from or otherwise corresponding to molecules found in species other than the human.

Chemical and functional equivalents of the subject nucleic acid or protein molecules should be understood as molecules exhibiting any one or more of the functional activities of these molecules and may be derived from any source such as being chemically synthesized or identified via screening processes such as natural product screening.

The derivatives include fragments having particular epitopes or parts of the entire protein fused to peptides, polypeptides or other proteinaceous or non-proteinaceous molecules.

Analogues contemplated herein include, but are not limited to, modification to side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, adenomaeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecules or their analogues.

Derivatives of nucleic acid sequences may similarly be derived from single or multiple nucleotide substitutions, deletions and/or additions including fusion with other nucleic acid molecules. The derivatives of the nucleic acid molecules of the present invention include oligonucleotides, PCR primers, antisense molecules, molecules suitable for use in cosuppression and fusion of nucleic acid molecules. Derivatives of nucleic acid sequences also include degenerate variants.

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with $NaBH_4$; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS);

acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with $NaBH_4$.

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivatisation, for example, to a corresponding amide.

Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carboethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during protein synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl

alanine and/or D-isomers of amino acids. A list of unnatural amino acids contemplated herein is shown in the following Table 1 below:

Non-conventional amino acid	Code	Non-conventional amino acid	Code
α-aminobutyric acid	Abu	L-N-methylalanine	Nmala
α-amino-α-methylbutyrate	Mgabu	L-N-methylarginine	Nmarg
aminocyclopropane-carboxylate	Cpro	L-N-methylasparagine	Nmasn
		L-N-methylaspartic acid	Nmasp
aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbornyl-carboxylate	Norb	L-N-methylglutamine	Nmgln
		L-N-methylglutamic acid	Nmglu
cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
cyclopentylalanine	Cpen	L-N-methylisolleucine	Nmile
D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
D-lysine	Dlys	L-N-methylthreonine	Nmthr
D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
D-threonine	Dthr	L-norleucine	Nle
D-tryptophan	Dtrp	L-norvaline	Nva
D-tyrosine	Dtyr	α-methyl-aminoisobutyrate	Maib

	D-valine	Dval	α-methyl- -aminobutyrate	Mgabu
	D-α-methylalanine	Dmala	α-methylcyclohexylalanine	Mchexa
	D-α-methylarginine	Dmarg	α-methylcylcopentylalanine	Mcpen
	D-α-methylasparagine	Dmasn	α-methyl-α-napthylalanine	Manap
	D-α-methylaspartate	Dmasp	α-methylpenicillamine	Mpen
	D-α-methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D-α-methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
	D-α-methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
	D-α-methylisoleucine	Dmile	N-amino-α-methylbutyrate	Nmaabu
	D-α-methylleucine	Dmleu	α-napthylalanine	Anap
	D-α-methyllysine	Dmlys	N-benzylglycine	Nphe
	D-α-methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	NgIn
	D-α-methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
	D-α-methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D-α-methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D-α-methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D-α-methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
	D-α-methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
	D-α-methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
	D-α-methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl))glycine	Nser
	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl))glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp
	D-N-methyllysine	Dnmlys	N-methyl-γ-aminobutyrate	Nmgabu
	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen

	N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
5	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
	D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	γ-aminobutyric acid	Gabu	N-(*p*-hydroxyphenyl)glycine	Nhtyr
	L-*t*-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
10	L-ethylglycine	Etg	penicillamine	Pen
	L-homophenylalanine	Hphe	L-α-methylalanine	Mala
	L-α-methylarginine	Marg	L-α-methylasparagine	Masn
	L-α-methylaspartate	Masp	L-α-methyl-*t*-butylglycine	Mtbug
	L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
15	L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
	L-α-methylhistidine	Mhis	L-α-methylhomophenylalanine	Mhphe
	L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
	L-α-methylleucine	Mleu	L-α-methyllysine	Mlys
	L-α-methylmethionine	Mmet	L-α-methylnorleucine	Mnle
20	L-α-methylnorvaline	Mnva	L-α-methylornithine	Morn
	L-α-methylphenylalanine	Mphe	L-α-methylproline	Mpro
	L-α-methylserine	Mser	L-α-methylthreonine	Mthr
	L-α-methyltryptophan	Mtrp	L-α-methyltyrosine	Mtyr
	L-α-methylvaline	Mval	L-N-methylhomophenylalanin	Nmhphe
25	N-(N-(2,2-diphenylethyl) carbamylmethyl)glycine	Nnbhm	N-(N-(3,3-diphenylpropyl) carbamylmethyl)glycine	Nnbhe
	1-carboxy-1-(2,2-diphenyl-ethylamino)cyclopropane	Nmbc		

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with n=1 to n=6, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional

reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety.

A further aspect of the present invention relates to the use of the invention in relation to the treatment and/or prophylaxis of disease conditions. Without limiting the present invention to any one theory or mode of action, the inventors have determined that increased levels of any one or more the *adenoma markers* (both novel and those which have been previously characterised but not recognised as being an adenoma marker) disclosed herein is associated with the onset or a predisposition to the onset of a neoplasm, in particular an adenoma. Accordingly, modulation of the expression and/or functional activity of these adenoma markers provides a mechanism for treating conditions characterised by aberrant, unwanted or otherwise inappropriate cell growth. Although the preferred method is to down-regulate uncontrolled cellular proliferation in an individual, by down-regulating the expression and/or functional activity of one or more of the adenoma markers disclosed herein, up-regulation of cell growth may also be desirable in certain circumstances such as to promote wound healing and angiogenesis or other human processes.

The present invention therefore contemplates a method for the treatment and/or prophylaxis of a condition characterised by aberrant, unwanted or otherwise inappropriate cell growth in a subject, said method comprising administering to said subject an effective amount of an agent for a time and under conditions sufficient to modulate *adenoma marker* expression and/or adenoma marker functional activity.

Reference to "aberrant, unwanted or otherwise inappropriate" cell growth should be understood as a reference to overactive cell growth, to physiologically normal cell growth which is inappropriate in that it is unwanted or to insufficient cell growth. Preferably said inappropriate cell growth is uncontrolled cell proliferation.

According to this preferred embodiment there is provided a method for the treatment and/or prophylaxis of a condition characterised by uncontrolled cell proliferation in a subject, said method comprising administering to said subject an effective amount of an

agent for a time and under conditions sufficient to down-regulate *adenoma marker* expression and/or adenoma marker functional activity.

Preferably said condition is a neoplastic condition and still more preferably an adenoma.

In a most preferred embodiment there is provided a method for the treatment and/or prophylaxis of an adenoma in a subject said method comprising administering to said subject an effective amount of and agent for a time and under conditions sufficient to down-regulate *adenoma marker* expression and/or adenoma marker functional activity.

Still more preferably said adenoma is a colorectal adenoma.

The method of the present invention preferably facilitates the subject proliferation being reduced, retarded or otherwise inhibited. Reference to "reduced, retarded or otherwise inhibited" should be understood as a reference to inducing or facilitating the partial or complete inhibition of cell proliferation. Said inhibition may occur either by direct or indirect mechanisms and includes the induction of cellular apoptosis or other cellular mechanisms.

An "effective amount" means an amount necessary at least partly to attain the desired response, or to delay the onset or inhibit progression or halt altogether, the onset or progression of a particular condition being treated. The amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated, the degree of protection desired, the formulation of the composition, the assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

Reference herein to "treatment" and "prophylaxis" is to be considered in its broadest context. The term "treatment" does not necessarily imply that a subject is treated until total recovery. Similarly, "prophylaxis" does not necessarily mean that the subject will not

eventually contract a disease condition. Accordingly, treatment and prophylaxis include amelioration of the symptoms of a particular condition or preventing or otherwise reducing the risk of developing a particular condition. The term "prophylaxis" may be considered as reducing the severity or onset of a particular condition. "Treatment" may also reduce the severity of an existing condition.

Administration of the agent (including an adenoma marker or functional equivalent, derivative, homologue, analogue or mimetic thereof or an *adenoma marker* nucleic acid molecule or derivative, equivalent, homologue or analogue thereof) [herein referred to as "modulatory agent"], in the form of a pharmaceutical composition, may be performed by any convenient means. The modulatory agent of the pharmaceutical composition is contemplated to exhibit therapeutic activity when administered in an amount which depends on the particular case. The variation depends, for example, on the human or animal and the modulatory agent chosen. A broad range of doses may be applicable. Considering a patient, for example, from about 0.1 mg to about 1 mg of modulatory agent may be administered per kilogram of body weight per day. Dosage regimes may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily, weekly, monthly or other suitable time intervals or the dose may be proportionally reduced as indicated by the exigencies of the situation.

The modulatory agent may be administered in a convenient manner such as by the oral, intravenous (where water soluble), intraperitoneal, intramuscular, subcutaneous, intradermal or suppository routes or implanting (e.g. using slow release molecules). The modulatory agent may be administered in the form of pharmaceutically acceptable non-toxic salts, such as acid addition salts or metal complexes, e.g. with zinc, iron or the like (which are considered as salts for purposes of this application). Illustrative of such acid addition salts are hydrochloride, hydrobromide, sulphate, phosphate, maleate, acetate, citrate, benzoate, succinate, malate, ascorbate, tartrate and the like. If the active ingredient is to be administered in tablet form, the tablet may contain a binder such as tragacanth, corn starch or gelatin; a disintegrating agent, such as alginic acid; and a lubricant, such as magnesium stearate.

Routes of administration include, but are not limited to, respiratorally, intratracheally, nasopharyngeally, intravenously, intraperitoneally, subcutaneously, intracranially, intradermally, intramuscularly, intraoccularly, intrathecally, intracereberally, intranasally, infusion, orally, rectally, *via* IV drip, patch and implant. Preferably, said route of administration is oral.

In accordance with these methods, the agent defined in accordance with the present invention may be coadministered with one or more other compounds or molecules. By "coadministered" is meant simultaneous administration in the same formulation or in two different formulations via the same or different routes or sequential administration by the same or different routes. By "sequential" administration is meant a time difference of from seconds, minutes, hours or days between the administration of the two types of molecules. These molecules may be administered in any order.

Another aspect of the present invention contemplates the use of an agent as hereinbefore defined in the manufacture of a medicament for the treatment of a condition in a mammal, which condition is characterised by the aberrant, unwanted or otherwise inappropriate cell growth wherein said agent modulates adenoma marker functional activity or *adenoma marker* expression.

Preferably said condition is a neoplastic condition and even more preferably an adenoma. Still more preferably, said adenoma is a colorectal adenoma.

In yet another further aspect, the present invention contemplates a pharmaceutical composition comprising a modulatory agent as hereinbefore defined and one or more pharmaceutically acceptable carriers and/or diluents. Said modulatory agents are referred to as the active ingredients.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous

preparation of sterile injectable solutions or dispersion or may be in the form of a cream or other form suitable for topical application. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of superfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilisation. Generally, dispersions are prepared by incorporating the various sterilised active ingredient into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible

tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions in such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1 μg and 2000 mg of active compound.

The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter: a binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such as sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound(s) may be incorporated into sustained-release preparations and formulations. The pharmaceutical composition may also comprise genetic molecules such as a vector capable of transfecting target cells where the vector carries a nucleic acid molecule encoding a modulatory agent. The vector may, for example, be a viral vector.

Yet another aspect of the present invention relates to modulatory agents, as hereinbefore defined, when used in the method of the present invention.

Still another aspect of the present invention is directed to antibodies to adenoma markers or *adenoma markers* including catalytic antibodies. Such antibodies may be monoclonal or polyclonal and may be selected from naturally occurring antibodies to adenoma markers or may be specifically raised to an adenoma marker. In the case of the latter, the adenoma marker may first need to be associated with a carrier molecule. The antibodies and/or recombinant adenoma marker of the present invention are particularly useful as therapeutic or diagnostic agents. Alternatively, fragments of antibodies may be used such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies. The antibodies of this aspect of the present invention are particularly useful for immunotherapy and may also be used as a diagnostic tool for assessing apoptosis or monitoring the program of a therapeutic regime.

For example, an adenoma marker can be used to screen for naturally occurring antibodies to an adenoma marker.

For example, specific antibodies can be used to screen for adenoma marker proteins. The latter would be important, for example, as a means for screening for levels of an adenoma marker in a cell extract or other biological fluid or purifying an adenoma marker made by recombinant means from culture supernatant fluid. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays, ELISA and flow cytometry. It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of an adenoma marker.

Both polyclonal and monoclonal antibodies are obtainable by immunization with the protein or peptide derivatives and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred

but are relatively easily prepared by injection of a suitable laboratory animal with an effective amount of an adenoma marker, or antigenic parts thereof, collecting serum from the animal, and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art. (See, for example Douillard and Hoffman (1981).

The present invention is further described by the following non-limiting examples:

SEQUENCE LISTING SUMMARY

Table 2

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
1	11-10a	*Adenoma marker* cDNA sequence	gi|15431275|gb|AC079240.6|Homo sapiens BAC clone RP11-808H.523 e-146
2	11-10e; and 11-5b	Claudin 2 cDNA sequence	Genbank Accession Number NM020384
3	11-10e; and 11-5b	Clauding 2 protein sequence	Genbank Accession Number NP065117
4	1-1d	*Adenoma marker* cDNA sequence	gi|5763746|emb|AL049766.14|HSDJ686N3 Human DNA sequence fro.. 373 e-100
5	1-1g	*Adenoma marker* cDNA sequence	gi|5763746|emb|AL049766.14|HSDJ686N3 Human DNA sequence from clone RP4-686N3 on chromosome 20q13.2-13.2 Contains the 3' part of the gene for a novel ATP dependent RNA helicase (contains conserved C-terminal helicase domains and DEAD-DEAH boxes), the KIAA1404 gene, a putative novel gene
6	12-17a	*Adenoma marker* cDNA sequence	gi|21629406|gb|AC099845.2|Homo sapiens chromosome 18, clon...456 e-126
7	12-2f 8-2d	*Adenoma marker* cDNA sequence	gi|18104869|gb|AC023302.9|Homo sapiens chromosome 15, clon.. 593 e-167
8	2-12e clone 8	*Adenoma marker* cDNA sequence	gi|18645166|gb|BC023990.1| Homo sapiens, annexin A2, clone ... 141 9e-31
9	2-13a clone 5	*Adenoma marker* cDNA sequence	
10	2-20b clone 2	*Adenoma marker* cDNA sequence	gi|21732430|emb|AL831917.1|HSM803250 Homo sapiens mRNA; cDNA DKFZp761F0118 (from clone DKFZp761F0118)
11	3-10e clone 6	*Adenoma marker* cDNA sequence	gi|15823776|dbj|AB063285.1|Homo sapiens HLCS gene for holocarboxylase synthetase, complete cds
12	3-12a	*Adenoma marker* cDNA sequence	gi|11597162|gb|AC013410.5|AC013410 Homo sapiens BAC clone RP11-495I2 from 2, complete sequence
13	3-16b clone 4	*Adenoma marker* cDNA sequence	gi|6382477|gb|AC005881.3|AC005881 citb_79_e_16, complete sequence [Homo sapiens]

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
14	4-14b	*Adenoma marker* cDNA sequence	
15	4-17d	*Adenoma marker* cDNA sequence	gi|18426891|gb|AC000385.2| Homo sapiens chromosome 11 clone PAC pDJ392a17, complete sequence i|21732802|emb|AL832255.1|HSM803562 Homo sapiens mRNA; cDNA DKFZp667D1717 (from clone DKFZp667D1717)
16	4-18e	*Adenoma marker* cDNA sequence	gi|17985585|gb|AF381996.1|AF381996 Homo sapiens haplotype M12 mitochondrion, complete genome
17	4-2a	*Adenoma marker* cDNA sequence	gi|17149447|gb|AC096915.2| Homo sapiens chromosome 3 clone RP11-24H1, complete sequence
18	5-2f	*Adenoma marker* cDNA sequence	gi|21359956|ref|NM_024569.2| Homo sapiens hypothetical protein FLJ21047 (FLJ21047), mRNA
19	5-2g	*Adenoma marker* cDNA sequence	gi|21749695|dbj|AK091346.1| Homo sapiens cDNA FLJ34027 fis, clone FCBBF2003549, highly similar to GAP-associated tyrosine phosphoprotein p62 (Sam68)
20	6-12a	*Adenoma marker* cDNA sequence	
21	6-12b	*Adenoma marker* cDNA sequence	
22	6-16c clone 1	*Adenoma marker* cDNA sequence	gi|2429080|dbj|D87675.1| Homo sapiens DNA for amyloid precursor protein, complete cds
23	6-17b	*Adenoma marker* cDNA sequence	gi|4502264|ref|NM_001675.1| Homo sapiens activating transcription factor 4 (tax-responsive enhancer element B67) (ATF4), mRNA
24	6-18d clone 7	*Adenoma marker* cDNA sequence	gi|19335766|emb|AL356738.14| Human DNA sequence from clone RP13-228J13 on chromosome X, complete sequence [Homo sapiens
25	7-12a	*Adenoma marker* cDNA sequence	gi|15991856|gb|BC012895.1|BC012895 Homo sapiens, clone MGC:18288 IMAGE:4179238, mRNA, complete cds
26	8-17a	*Adenoma marker* cDNA sequence	gi|3873300|gb|AC005829.1|AC005829 Homo sapiens chromosome 17, clone hRPK.259_G_18, complete sequence
27	8-19a	*Adenoma marker* cDNA sequence transcript 1	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
28	8-19a	*Adenoma marker* cDNA sequence transcript 2	
29	8-19a	*Adenoma marker* cDNA sequence transcript 3	
30	8-7bi	*Adenoma marker* cDNA sequence	
31	9-4g clone 5	*Adenoma marker* cDNA sequence	gi|21700762|ref|NM_144570.1| Homo sapiens HN1 like (HN1L), mRNA
32	9-8a; and 2-20a-2	Transposon L1.1 cDNA sequence	Genbank Accession Number M80340
33	9-8a; and 2-20a-2	Transposon L1.1 CDS 1 protein sequence	Genbank Accession Number AAA51621
34	9-8a; and 2-20a-2	Transposon L1.1 CDS 2 protein sequence	Genbank Accession Number AAA51622
35	9-8f2 clone 5	*Adenoma marker* cDNA sequence	gi|6562085|emb|AL078591.18|HSDJ198I9 Human DNA sequence from clone RP1-198I9 on chromosome 6q12-13. Contains the gene KIAA1411, ESTs, STSs and GSSs, complete sequence
36	9-8g	*Adenoma marker* cDNA sequence	gi|21733549|emb|AL832961.1|HSM804272 Homo sapiens mRNA; cDNA DKFZp666O0110 (from clone DKFZp666O0110)
37	9-8j2 clone 4	*Adenoma marker* cDNA sequence	gi|22002110|gb|AC022080.37| Homo sapiens 12 BAC RP11-820K3 (Roswell Park Cancer Institute Human BAC
38	11-10b; 3-2c; 12-7c; 9-2d; and 11-2d	Regenerating Protein IV cDNA sequence	Genbank Accession Number NM032044
39	11-10b; 3-2c; 12-7c; 9-2d; and 11-2d	Regenerating Protein IV protein sequence	Genbank Accession Number NP114433
40	11-20e	*Adenoma marker* cDNA sequence	gi|20539919|ref|XM_168302.1| Homo sapiens zinc finger prote... 854 0.0
41	3-19e	*Adenoma marker* cDNA sequence	gi|20539919|ref|XM_168302.1| Homo sapiens zinc finger protein 36 (KOX 18) (ZNF36), mRNA
42	1-19e	*Adenoma marker* cDNA sequence	gi|16933566|ref|NM_005370.3| Homo sapiens mel transforming ... 884 0.0

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
43	1-6aii	Gastric Intrinsic Factor cDNA sequence	Genbank Accession Number NM005142
44	1-6aii	Gastric Intrinsic Factor protein sequence	Genbank Accession Number NP005133
45	2-10b	*Adenoma marker* cDNA sequence	gi|6633800|ref|NM_007329.1| Homo sapiens deleted in maligna... 404 e-110
46	2-12f	*Adenoma marker* cDNA sequence	gi|21166384|ref|NM_138737.1| Homo sapiens hephaestin (HEPH)... 791 0.0
47	3-13e	*Adenoma marker* cDNA sequence	
48	2-18f clone 5	*Adenoma marker* cDNA sequence	gi|5051939|gb|AF143313.1|PTEN2 Homo sapiens PTEN (PTEN) gen... 569 e-160
49	2-1c	Defensin alpha 6 cDNA sequence	Genbank Accession Number NM001926
50	2-1c	Defensin alpha 6 protein sequence	Genbank Accession Number NP001917
51	2-1g	*Adenoma marker* cDNA sequence	gi|19923767|ref|NM_005682.2| Homo sapiens G protein-coupled receptor 56 (GPR56), mRNA
52	2-7g clone 4	*Adenoma marker* cDNA sequence	Genbank Accession Number NM 022149.1
53	3-12e clone 3	*Adenoma marker* cDNA sequence	Accession Number XM 010264.2
54	3-16k	*Adenoma marker* cDNA sequence	gi|12654696|gb|BC001188.1|BC001188 Homo sapiens, transferri... 523 e-146
55	3-5c clone 4	*Adenoma marker* cDNA sequence	gi|4758719|ref|NM_004529.1| Homo sapiens myeloid-lymphoid or mixed-lineage leukemia
56	4-11e; and 5-13d	Transforming growth factor beta cDNA sequence	Genbank Accession Number NM000358
57	4-11e; and 5-13d	Transforming growth factor beta protein sequence	Genbank Accession Number NP000349
58	4-16d	*Adenoma marker* cDNA sequence	gi|17985823|gb|AF382013.1|AF382013 Homo sapiens haplotype M*2 mitochondrion, complete genome Length = 16567
59	4-18d	*Adenoma marker* cDNA sequence	Accession Number XM 015882.1

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
60	5-14j	GW112 cDNA sequence	Genbank Accession Number NM006418
61	5-14j	GW112 protein sequence	Genbank Accession Number NP006409
62	5-4a	S100 calcium binding protein P cDNA sequence	Genbank Accession Number NM005980
63	5-4a	S100 calcium binding protein P protein sequence	Genbank Accession Number NP005971
64	6-10d	*Adenoma marker* cDNA sequence	gi|20127471|ref|NM_005239.2| Homo sapiens v-ets erythroblastosis virus E26 oncogene homolog
65	6-16a	*Adenoma marker* cDNA sequence	gi|15145625|gb|AC023150.5|Homo sapiens BAC clone RP11-709L9 from 4, complete sequence
66	7-13b	SLC12A2 cDNA sequence	Genbank Accession Number NM001046
67	7-13b	SLC12A2 protein sequence	Genbank Accession Number NP001037
68	7-13d clone 4	*Adenoma marker* cDNA sequence	Accession Number M22146.1
69	8-12b	*Adenoma marker* cDNA sequence	Accession Number gbU16738.1
70	5-12a	*Adenoma marker* cDNA sequence	Accession Number gbU16738.1
71	8-16b	*Adenoma marker* cDNA sequence	|7739724|gb|AF257305.1|AF257305 Homo sapiens ASH1 mRNA, complete cds
72	9-13c3	*Adenoma marker* cDNA sequence	gi|21707883|gb|BC034141.1| Homo sapiens, similar to anti TNF-alpha antibody light-chain Fab >gi|21707883|gb|BC034141.1| Homo sapiens, similar to anti TNF-alpha antibody light-chain
73	3-1d2	*Adenoma marker* cDNA sequence	
74	3-1 cl 4	*Adenoma marker* cDNA sequence	
75	3-1 cl 7	*Adenoma marker* cDNA sequence	
76	4-1 cl 11	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
77	4-1 cl 13	*Adenoma marker* cDNA sequence	
78	4-1 cl 17	*Adenoma marker* cDNA sequence	
79	6-1a	*Adenoma marker* cDNA sequence	
80	6-3d	*Adenoma marker* cDNA sequence	
81	2-6a	*Adenoma marker* cDNA sequence	
82	2-7n	*Adenoma marker* cDNA sequence	
83	2-8r	*Adenoma marker* cDNA sequence	
84	6-5d	*Adenoma marker* cDNA sequence	
85	7-16g	*Adenoma marker* cDNA sequence	
86	7-17b2	*Adenoma marker* cDNA sequence	
87	8-2a	*Adenoma marker* cDNA sequence	
88	8-2c	*Adenoma marker* cDNA sequence	
89	8-12d	*Adenoma marker* cDNA sequence	
90	8-15a	*Adenoma marker* cDNA sequence	
91	9-8h	*Adenoma marker* cDNA sequence	
92	9-8i	*Adenoma marker* cDNA sequence	
93	9-10e	*Adenoma marker* cDNA sequence	
94	9-12a	*Adenoma marker* cDNA sequence	
95	9-14a	*Adenoma marker* cDNA sequence	
96	9-15a	*Adenoma marker* cDNA sequence	
97	9-16h	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
98	9-17a	*Adenoma marker* cDNA sequence	
99	9-19b	*Adenoma marker* cDNA sequence	
100	9-19c	*Adenoma marker* cDNA sequence	
101	10-7b	*Adenoma marker* cDNA sequence	
102	10-8a	*Adenoma marker* cDNA sequence	
103	10-8c	*Adenoma marker* cDNA sequence	
104	10-11b	*Adenoma marker* cDNA sequence	
105	10-12a	*Adenoma marker* cDNA sequence	
106	10-14j	*Adenoma marker* cDNA sequence	
107	10-16a	*Adenoma marker* cDNA sequence	
108	10-17b	*Adenoma marker* cDNA sequence	
109	10-17c	*Adenoma marker* cDNA sequence	
110	11-12d	*Adenoma marker* cDNA sequence	
111	11-12a	*Adenoma marker* cDNA sequence	
112	11-17a	*Adenoma marker* cDNA sequence	
113	11-17c	*Adenoma marker* cDNA sequence	
114	11-17d	*Adenoma marker* cDNA sequence	
115	11-19i	*Adenoma marker* cDNA sequence	
116	12-20b	*Adenoma marker* cDNA sequence	
117	12-7a	*Adenoma marker* cDNA sequence	
118	12-7d	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
119	1-1B-1	*Adenoma marker* cDNA sequence	
120	1-1H	*Adenoma marker* cDNA sequence	
121	1-11B2	*Adenoma marker* cDNA sequence	
122	1-14A	*Adenoma marker* cDNA sequence	
123	1-14B	*Adenoma marker* cDNA sequence	
124	1-14E2	*Adenoma marker* cDNA sequence	
125	1-15D	*Adenoma marker* cDNA sequence	
126	1-16A-3	*Adenoma marker* cDNA sequence	
127	1-16	*Adenoma marker* cDNA sequence	
128	1-16A-6	*Adenoma marker* cDNA sequence	
129	1-20B	*Adenoma marker* cDNA sequence	
130	1-20C	*Adenoma marker* cDNA sequence	
131	2-8A	*Adenoma marker* cDNA sequence	
132	2-8R	*Adenoma marker* cDNA sequence	
133	2-13A-2	*Adenoma marker* cDNA sequence	
134	2-17A-1	*Adenoma marker* cDNA sequence	
135	2-17A-5	*Adenoma marker* cDNA sequence	
136	2-17A-6	*Adenoma marker* cDNA sequence	
137	2-17F	*Adenoma marker* cDNA sequence	
138	2-17I-2	*Adenoma marker* cDNA sequence	
139	2-18D-1	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
140	2-18D-5	*Adenoma marker* cDNA sequence	
141	2-18F-1	*Adenoma marker* cDNA sequence	
142	2-19A	*Adenoma marker* cDNA sequence	
143	2-19C	*Adenoma marker* cDNA sequence	
144	2-20B-1	*Adenoma marker* cDNA sequence	
145	2-20C	*Adenoma marker* cDNA sequence	
146	2-20D	*Adenoma marker* cDNA sequence	
147	3-3F-6	*Adenoma marker* cDNA sequence	
148	3-5C-2	*Adenoma marker* cDNA sequence	
149	3-5C-5	*Adenoma marker* cDNA sequence	
150	3-5D-2	*Adenoma marker* cDNA sequence	
151	3-8A-3	*Adenoma marker* cDNA sequence	
152	3-8A-5	*Adenoma marker* cDNA sequence	
153	3-8A-7	*Adenoma marker* cDNA sequence	
154	3-8E-5	*Adenoma marker* cDNA sequence	
155	3-10A-5	*Adenoma marker* cDNA sequence	
156	3-10A-8	*Adenoma marker* cDNA sequence	
157	3-10B	*Adenoma marker* cDNA sequence	
158	3-11A	*Adenoma marker* cDNA sequence	
159	3-11G	*Adenoma marker* cDNA sequence	
160	3-14E	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
161	3-16H-6	*Adenoma marker* cDNA sequence	
162	3-16J-1	*Adenoma marker* cDNA sequence	
163	4-17C-3	*Adenoma marker* cDNA sequence	
164	4-18B	*Adenoma marker* cDNA sequence	
165	5-1C-1	*Adenoma marker* cDNA sequence	
166	5-1C-2	*Adenoma marker* cDNA sequence	
167	5-1E-4	*Adenoma marker* cDNA sequence	
168	5-10A	*Adenoma marker* cDNA sequence	
169	5-14M	*Adenoma marker* cDNA sequence	
170	5-14N-1	*Adenoma marker* cDNA sequence	
171	5-15C	*Adenoma marker* cDNA sequence	
172	5-16A	*Adenoma marker* cDNA sequence	
173	5-16C	*Adenoma marker* cDNA sequence	
174	5-17C	*Adenoma marker* cDNA sequence	
175	5-17D	*Adenoma marker* cDNA sequence	
176	5-19A	*Adenoma marker* cDNA sequence	
177	5-19H	*Adenoma marker* cDNA sequence	
178	5-20D	*Adenoma marker* cDNA sequence	
179	6-3D	*Adenoma marker* cDNA sequence	
180	6-6B2-1	*Adenoma marker* cDNA sequence	
181	6-14A	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
182	6-14B	*Adenoma marker* cDNA sequence	
183	6-17H-3	*Adenoma marker* cDNA sequence	
184	6-18B	*Adenoma marker* cDNA sequence	
185	6-18F	*Adenoma marker* cDNA sequence	
186	6-19AIII	*Adenoma marker* cDNA sequence	
187	6-20E-1	*Adenoma marker* cDNA sequence	
188	6-20E-3	*Adenoma marker* cDNA sequence	
189	7-1D-1	*Adenoma marker* cDNA sequence	
190	7-7D-1	*Adenoma marker* cDNA sequence	
191	7-11B2-3	*Adenoma marker* cDNA sequence	
192	7-12C	*Adenoma marker* cDNA sequence	
193	7-16E	*Adenoma marker* cDNA sequence	
194	7-17D	*Adenoma marker* cDNA sequence	
195	7-18C-2	*Adenoma marker* cDNA sequence	
196	7-19I-2	*Adenoma marker* cDNA sequence	
197	7-19I-5	*Adenoma marker* cDNA sequence	
198	7-19I-6	*Adenoma marker* cDNA sequence	
199	8-1A-2	*Adenoma marker* cDNA sequence	
200	8-5A	*Adenoma marker* cDNA sequence	
201	8-5D	*Adenoma marker* cDNA sequence	
202	8-5E	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
203	8-13E-2	*Adenoma marker* cDNA sequence	
204	8-16E	*Adenoma marker* cDNA sequence	
205	8-17C	*Adenoma marker* cDNA sequence	
206	8-19D	*Adenoma marker* cDNA sequence	
207	8-20A	*Adenoma marker* cDNA sequence	
208	8-1B-1	*Adenoma marker* cDNA sequence	
209	9-2B	*Adenoma marker* cDNA sequence	
210	9-7A-1	*Adenoma marker* cDNA sequence	
211	9-8F2-2	*Adenoma marker* cDNA sequence	
212	9-12C-5	*Adenoma marker* cDNA sequence	
213	9-16E-3	*Adenoma marker* cDNA sequence	
214	9-17A-2+5	*Adenoma marker* cDNA sequence	
215	9-17B	*Adenoma marker* cDNA sequence	
216	9-17D-2	*Adenoma marker* cDNA sequence	
217	10-4B-6	*Adenoma marker* cDNA sequence	
218	12-6D	*Adenoma marker* cDNA sequence	
219	12-17B	*Adenoma marker* cDNA sequence	
220	4-1 cl 2	*Adenoma marker* cDNA sequence	
221	4-1 cl 8	*Adenoma marker* cDNA sequence	
222	5-2c	*Adenoma marker* cDNA sequence	
223	1-10a	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
224	2-1e	*Adenoma marker* cDNA sequence	
225	2-6f	*Adenoma marker* cDNA sequence	
226	2-6i	*Adenoma marker* cDNA sequence	
227	2-7a	*Adenoma marker* cDNA sequence	
228	2-7b	*Adenoma marker* cDNA sequence	
229	2-7d	*Adenoma marker* cDNA sequence	
230	2-8f	*Adenoma marker* cDNA sequence	
231	5-1a	*Adenoma marker* cDNA sequence	
232	6-2a	*Adenoma marker* cDNA sequence	
233	6-5c	*Adenoma marker* cDNA sequence	
234	6-6d	*Adenoma marker* cDNA sequence	
235	7-5a	*Adenoma marker* cDNA sequence	
236	7-10b	*Adenoma marker* cDNA sequence	
237	7-11d3	*Adenoma marker* cDNA sequence	
238	7-18a	*Adenoma marker* cDNA sequence	
239	7-20b	*Adenoma marker* cDNA sequence	
240	8-2f	*Adenoma marker* cDNA sequence	
241	9-10c	*Adenoma marker* cDNA sequence	
242	9-14c	*Adenoma marker* cDNA sequence	
243	9-14g	*Adenoma marker* cDNA sequence	
244	9-16a	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
245	9-16b	*Adenoma marker* cDNA sequence	
246	9-19a	*Adenoma marker* cDNA sequence	
247	9-20a	*Adenoma marker* cDNA sequence	
248	10-10a	*Adenoma marker* cDNA sequence	
249	11-11b	*Adenoma marker* cDNA sequence	
250	5-13E	*Adenoma marker* cDNA sequence	
251	2-13B	*Adenoma marker* cDNA sequence	
252	11-12c	*Adenoma marker* cDNA sequence	
253	11-13a	*Adenoma marker* cDNA sequence	
254	11-13d	*Adenoma marker* cDNA sequence	
255	11-20b	*Adenoma marker* cDNA sequence	
256	11-20d	*Adenoma marker* cDNA sequence	
257	11-5f	*Adenoma marker* cDNA sequence	
258	12-13b	*Adenoma marker* cDNA sequence	
259	12-15a	*Adenoma marker* cDNA sequence	
260	1-1B-4	*Adenoma marker* cDNA sequence	
261	1-7F-2	*Adenoma marker* cDNA sequence	
262	1-7G-2	*Adenoma marker* cDNA sequence	
263	1-8E	*Adenoma marker* cDNA sequence	
264	1-9B	*Adenoma marker* cDNA sequence	
265	2-5G	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
266	2-6G-1	*Adenoma marker* cDNA sequence	
267	2-6G-2	*Adenoma marker* cDNA sequence	
268	2-7G-6	*Adenoma marker* cDNA sequence	
269	2-8G-1	*Adenoma marker* cDNA sequence	
270	2-8G-2	*Adenoma marker* cDNA sequence	
271	2-8Q	*Adenoma marker* cDNA sequence	
272	2-11A	*Adenoma marker* cDNA sequence	
273	2-12D	*Adenoma marker* cDNA sequence	
274	2-12E-7	*Adenoma marker* cDNA sequence	
275	2-17H	*Adenoma marker* cDNA sequence	
276	2-17I-5	*Adenoma marker* cDNA sequence	
277	2-17I-1	*Adenoma marker* cDNA sequence	
278	2-18C-2	*Adenoma marker* cDNA sequence	
279	2-18C-5	*Adenoma marker* cDNA sequence	
280	2-18G	*Adenoma marker* cDNA sequence	
281	3-2C-A	*Adenoma marker* cDNA sequence	
282	3-3A	*Adenoma marker* cDNA sequence	
283	3-3F-5	*Adenoma marker* cDNA sequence	
284	3-3F-7	*Adenoma marker* cDNA sequence	
285	3-5C-3	*Adenoma marker* cDNA sequence	
286	3-19J-1	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
287	5-18F-2	*Adenoma marker* cDNA sequence	
288	5-19I	*Adenoma marker* cDNA sequence	
289	3-5D-6	*Adenoma marker* cDNA sequence	
290	3-8B	*Adenoma marker* cDNA sequence	
291	3-10A-6	*Adenoma marker* cDNA sequence	
292	3-16B-3	*Adenoma marker* cDNA sequence	
293	3-16H-5	*Adenoma marker* cDNA sequence	
294	3-16HII-5	*Adenoma marker* cDNA sequence	
295	3-16J-2	*Adenoma marker* cDNA sequence	
296	3-18E-1	*Adenoma marker* cDNA sequence	
297	3-18E-6	*Adenoma marker* cDNA sequence	
298	4-16E	*Adenoma marker* cDNA sequence	
299	5-2I	*Adenoma marker* cDNA sequence	
300	5-5A	*Adenoma marker* cDNA sequence	
301	5-8B	*Adenoma marker* cDNA sequence	
302	5-14A	*Adenoma marker* cDNA sequence	
303	5-14N-2	*Adenoma marker* cDNA sequence	
304	5-15C	*Adenoma marker* cDNA sequence	
305	5-16B	*Adenoma marker* cDNA sequence	
306	5-17A	*Adenoma marker* cDNA sequence	
307	5-18F-6	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone-information
308	5-18F-7	*Adenoma marker* cDNA sequence	
309	5-20C-3	*Adenoma marker* cDNA sequence	
310	6-1F-4	*Adenoma marker* cDNA sequence	
311	6-3C-2	*Adenoma marker* cDNA sequence	
312	6-3C-4	*Adenoma marker* cDNA sequence	
313	6-5E-2	*Adenoma marker* cDNA sequence	
314	6-10C	*Adenoma marker* cDNA sequence	
315	6-12G	*Adenoma marker* cDNA sequence	
316	6-17B	*Adenoma marker* cDNA sequence	
317	6-20E-2	*Adenoma marker* cDNA sequence	
318	7-2D	*Adenoma marker* cDNA sequence	
319	7-11A	*Adenoma marker* cDNA sequence	
320	7-11C2	*Adenoma marker* cDNA sequence	
321	7-13D-2	*Adenoma marker* cDNA sequence	
322	7-20A-4	*Adenoma marker* cDNA sequence	
323	7-20D-2	*Adenoma marker* cDNA sequence	
324	7-20D-3	*Adenoma marker* cDNA sequence	
325	8-1D	*Adenoma marker* cDNA sequence	
326	8-5F4-1	*Adenoma marker* cDNA sequence	
327	8-13E-5	*Adenoma marker* cDNA sequence	
328	8-14C-1	*Adenoma marker* cDNA sequence	

PCT Sequence ID Number	*Adenoma Marker* Clone Name	Sequence Description	Additional clone information
329	9-1A-1	*Adenoma marker* cDNA sequence	
330	8-15B2	*Adenoma marker* cDNA sequence	
331	9-18A-2	*Adenoma marker* cDNA sequence	
332	10-4B-3	*Adenoma marker* cDNA sequence	
333	10-10D	*Adenoma marker* cDNA sequence	
334	11-12E	*Adenoma marker* cDNA sequence	
335	12-2C	*Adenoma marker* cDNA sequence	
336	XS81	*Adenoma marker* cDNA sequence	
337	2-5a	*Adenoma marker* cDNA sequence	
338	HumregA	*Adenoma marker* cDNA sequence	

EXAMPLE 1
ADENOMA COLLECTION AND RNA ISOLATION

a) Samples of adenoma and normal tissue were obtained from patients undergoing colonoscopy. A portion from each lesion was allocated for routine diagnostic pathology analysis and the remainder quick frozen and stored at –70°C.

b) RNA was extracted from frozen tissue using a standard guanidine thiocyanate, acid phenol method. The quality of each RNA preparation was examined by denaturing agarose gel electrophoresis.

Characterization Of Novel Genes

c) The adenoma cDNA library was prepared in a bacteriophage lambda vector (λSCREEN-1) using commercially available reagents obtained from Novagen. Library screening was performed under standard conditions at 65°C using 4xSSC hybridization buffer. Filters were washed twice at 65°C in 2xSSC for 30 minutes each and then at 65°C in 0.2xSSC for 15 minutes each. Filters were then exposed to X-ray film overnight at -70°C in the presence of one intensifying screen. Positive plaques were picked, grown up and rescreened under the same conditions until a homogeneous population emerged. The inserts from individual phage were then characterized by restriction enzyme mapping and DNA sequence analysis.

In some cases library screening was performed by PCR screening of pools of library clones using sequence information from the cloned differential display products to design appropriate primers. DNA fragments of the predicted size were identified after gel electrophoresis of the PCR reaction. These bands were excised from the gel, cloned and sequenced.

d) Sequence of the 5' and 3' ends of selected mRNAs was determined by the Rapid Amplification of cDNA Ends (RACE) technique using 5-10μg of total RNA. The

FirstChoice kit made by Ambion was used for these experiments according to the suppliers instructions.

e) Northern blot analysis was performed with 2μg of poly(A)+ RNA which was electrophoresed on a 1% agarose gel containing 1% formaldehyde. RNA was transferred from the gel to a Hybond N+ membrane (Amersham Pharmacia Biotech) which was then hybridized to selected probes under conditions described for library screening.

EXAMPLE 2

DIFFERENTIAL DISPLAY ANALYSIS OF COLONIC TISSUE RNA

1. Total RNA (DNAse treated) isolated from up to 32 patient samples was reverse transcribed with one of 12 anchored primers (AP1-12) from the set $T7dT_{12}VN$ to generate one of 12 cDNA samples for each patient sample to be used in subsequent differential display PCR.

   where:- T7 = the last 17bp of the T7 primer sequence;
   V = the bases A,C or G
   N = the bases A,C,G or T

2. An aliquot of each of up to 32 patient cDNA samples was used in subsequent PCR's (Using Applied Biosystems Amplitaq Gold) using the appropriate $T7dT_{12}VN$ primer with one of a series of 20 arbitrary primers comprising the last 16bp of an M13 primer sequence followed by a 10bp core annealing sequence.

3. PCR conditions led to 4 cycles of amplification based around core sequence annealing to cDNA template followed by 25 cycles of PCR in which amplicons generated in the first four rounds are exponentially amplified.

4. α^{32}P-dATP incorporation into PCR products allowed visualisation after they had been electrophoresed through a large format polyacrylamide gel at 850 volts overnight at 50°C.

5. Gels were then air dried and exposed to X-ray film at room temperature for 24-36 hours. Film images of gels were then examined for bands which were more intense in the adenoma samples as compared to normals. The film was then used as a template to locate the appropriate region on the dried gel and the required bands excised with a scalpel blade.

6. DNA was passively eluted from excised bands in 1 X TE solution overnight (TE = 10mMTris-HCl pH7.4, 1mM EDTA)

7. An aliquot of the eluate was then subjected to further rounds of PCR to generate enough material for cloning and DNA sequence analysis using Big Dye Terminator (Applied Biosystems) chemistry according to manufacturers instructions. .

EXAMPLE 3

REAL TIME PCR CONFIRMATION OF DIFFERENTIAL DISPLAY CLONES

1. Sequences isolated from differential display were compared using the BLAST algorithm to sequence databases housed on the National Centre for Biotechnology Information server located at http://www.ncbi.nlm.nih.gov/BLAST/.

Primer sets were then designed for selected sequences so that accurate measurement of tissue mRNA levels could be determined by quantitative PCR. Total RNA was reverse transcribed into cDNA using an oligo(dT) primer and Superscript II (Invitrogen) enzyme according to standard methods. Each cDNA population was then analyzed by real time PCR using a Corbett Research Rotorgene 2000 and reagents from a SYBR Green PCR Master Mix kit (Applied Biosystems). Cycle thresholds were then computed using Rotorgene 2000 version 4.6 software. The fold

elevation in the level of each mRNA was then calculated according to the formula $2^{(Nt-Nc)-(Tt-Tc)}$ where Nt is the cycle threshold for the test gene observed in the normal tissue, Nc is the cycle threshold for the control gene observed in the normal tissue, Tt is the cycle threshold for the test gene in the tumour and Tc is the cycle threshold for the control gene in the tumour. All calculations were performed assuming 100 percent efficiency at each PCR cycle. The control mRNA used to calculate Nc and Tc was β-actin.

EXAMPLE 4

mRNA EXPRESSION ANALYSIS FOR COLORECTAL ADENOMAS

Single Marker Analysis

(i) *Background*

67 mRNA sequences were isolated using differential display analysis. The expression level of the Markers was quantified using QRTPCR in 71 tissue samples (21 normal, 20 Tubular Adenoma, 26 Tubulovillous Adenoma, and 4 Villous Adenoma.) This expression data has been tabulated as "fold increase" in expression levels for each adenoma tissue over the mean expression level of Normal tissues, as previously described.

Two analytical approaches have been used to investigate the diagnostic utility of the Markers in Normal and Adenoma Tissues. First, we explored the upregulation for each Marker across the range of tissues in terms of total fold upregulation. Further, cluster analysis was used to assess the utility of the candidate Markers by identifying subsets of the 67 Markers that correctly discriminate between the Normal and Adenoma tissues. For the purposes of this analysis, "Adenoma" tissue includes all histological grades.

(ii) *Results*

To analyze the Markers individually we rank-ordered each Marker in terms of three criteria: a) average fold upregulation for Adenoma tissues relative to the average Normal

expression level, b) the percentage of disease tissues expressing greater than 5-fold upregulation, and c) the percentage of normal tissues expressing greater than 5-fold upregulation.

Average Upregulation

One diagnostic application using these Markers may be to establish a clinically relevant threshold of over-expression for one or more Markers relative to normal colorectal epithelium expression levels. Twenty-eight (28) Markers were determined to yield a five fold or greater average expression for Adenoma tissues relative to the average expression levels of the Normal tissues. A further twenty-six (26) markers were shown to express an average of 2 to 4 fold over the average normal signal. These Markers are listed in Table 3 and Table 4, respectively.

Percent Of Disease Tissues Demonstrating Upregulation

To assess the application "Sensitivity" using individual Markers, each Marker clone was ranked according to the percent of Adenoma tissues that would be identified for a given threshold of expression. Twenty-four (24) Markers were shown to express at least five-fold higher in 50% of individual tissues relative to the average Normal expression level. These Markers are listed in Table 5. Using multiple markers from this list in combination yields a higher apparent sensitivity in terms of the number of diseased tissues included and a higher specificity in terms of the percentage of normal tissues excluded.

Percent Of Normal Tissues Demonstrating Upregulation

To assess the "Specificity" for individual Markers, we re-evaluated the marker lists identified in Table 3 and Table 5 in light of expression levels also demonstrated for individual Normal Tissues. For this purpose Table 6 and Table 7 combine the sensitivity measures of Tables 3 and 5 with a threshold value of 5-fold over-expression in any individual Normal tissue. The Tables include all such markers that are upregulated greater

than 5 fold in less than 20% of Normal tissues. Twenty-five Markers were shown to yield over 5-fold expression in less than 20% of Normal tissues as shown in Table 6. Of these markers in Table 4, twenty-one (21) were also identified as expressing greater than 5-fold in at least 50% of Adenoma tissues while not over-expressing in greater than 20% of Normal tissues. These Markers are listed in Table 7.

EXAMPLE 5

mRNA EXPRESSION ANALYSIS FOR COLORECTAL ADENOMAS

Cluster Analysis

Cluster analysis showed that near perfect discrimination (70/71) can be achieved for one set of three markers and six unique sets of four markers.

(i) *Methods*

This cluster analysis is based on the k-nearest neighbor (KNN) technique described in Li, L., Darden, T., Weinberg, C., Levine, A. and Pedersen, L. (2001) Gene Assssment and Sample classification for gene expression data using a genetic algorithm/k nearest neighbor method. Combinatorial Chemistry & High Throughput Screening, Vol. 4(8), 727-739. whereby a given Tissue (X) is classified according to the class membership of the k tissues nearest to (X) in n-dimensional space described by expression levels of genes in the Marker set. Tissues were considered unclassified (and "missed") if analysis of the k-nearest neighbors tissues failed to achieve a unanimous result. In this analysis a range of k values was explored (1,2,...,5) and the final results are calculated based k=3. As expected, increasing k values result in lowers numbers of correctly classified tissues.

Successful classification is measured by comparison of the KNN derived tissue class against pathology diagnosis.

The array of expression values for each Marker in each Tissue forms the raw data for cluster analysis.

Diagnostically Useful Marker Sets

A set of markers is determined to be diagnostically useful if that set provides discrimination between tissue classes of interest (e.g. between Normal and Adenoma tissues) across a reasonably large sample of known tissues.

To illustrate this discrimination, it has been shown that the expression levels of two genes, designated clones 8/2d and 11-10a, provide segregation (or classification) between 21 Normal tissues and 50 Adenoma tissues that we have analyzed. Using these two markers, the k-nearest neighbor analysis is able to properly classify 68 out of the total 71 tissues. This classification is demonstrated as follows:

The average level of expression for each marker was calculated relative to the average level of its expression in normal tissue. Each value is then log normalized to give an expression table (Table 8) for the two markers such as:

Table 8

Num	Tissue ID	Type	Log (8/2d Fold ↑)	Log (11/10a Fold ↑)
1	A1	Norm	-0.3180	-1.7505
2	A2	Norm	0.0797	0.3718
3	A3	Norm	0.6337	0.5975
4	A4	Norm	0.4982	0.6487
5	A5	Norm	-0.2363	-0.0158

↓

69	I5	Adenoma	1.9401	0.4982
70	I6	Adenoma	1.8498	0.4440
71	I7	Adenoma	2.1840	0.3146

In this two-marker analysis, each tissue is therefore specified by the (two dimensional) coordinates described by the log values, e.g. Tissue A1 : (-0.0318, -1.7505), A2 : (0.0797,

0.3718),..., I7: (2.1840, 0.3146). Finally the distance is calculated between each tissue to determine which tissues are "nearest" to each other in terms of the chosen markers. For the two dimensional analysis using just the markers 8/2d and 11/10a, this relationship is conveniently visualized in a Cartesian plot of all tissue data points. This plot is shown in Figure 1.

Visual inspection of this graph clearly shows the relationship between each tissue, and demonstrates the self-clustering between the Normal tissues and Adenoma tissues.

To measure the utility of unique marker sets to provide class discrimination we have chosen the k-nearest neighbor (KNN) metric, where each data point (tissue) is classified according to its k nearest neighbors in the plot space according to the Euclidean distance formula:

A is a point described by the vector, $A = \{ \xi_1, \xi_2, \xi_3, \ldots \xi_n \}$, and

B is a point described by the vector, $B = \{ \eta_1, \eta_2, \eta_3, \ldots \eta_n \}$, then

$$\rho_E (A, B) = \| A - B \|$$

$$= +\sqrt{\sum_{i=1}^{n} (\xi_i - \eta_i)^2} \quad \text{(positive root only)}$$

To assure a robust clustering analysis we have chosen to use a k value of three (3). In other words, each tissue is calculated to belong to the same class as the three nearest data points in the plot. Next, the *calculated* tissue class is compared to the *known* tissue class as defined by histopathology to determine whether the KNN classification is correct for that tissue. If the three nearest tissues do not agree (i.e. one of the neighbors is of a different class to the other two) then the tissue is considered unclassified. Finally, the number of correctly classified tissues is totalled to provide a measure of classification strength for that particular set of markers.

In this plot, only three tissue data points, A7, C5, and H1 -- two presumed normal tissues and one adenoma, are misclassified using markers 8/2d and 11/10a with this KNN rule yielding a diagnostic accuracy of 96% (68/71). These points are highlighted in Figure 2.

Using this method, larger marker sets were explored in higher dimensions for clustering potential. By evaluating all possible combinations of three marker sets we have identified one set that is 99% (70/71) accurate. In four dimensions we have identified six unique sets that also achieve 99% accuracy. In all cases, one Normal tissue is missed (A7 in 5 cases, and C5 in 1 case) yielding an apparent sensitivity of 100% (50/50) and specificity of 95% (20/21) for these six sets.

It is instructive to note that this classification technique is ***not*** simply a reflection of the average fold upregulation for each marker in adenomas relative to normal tissues. In fact, while the first marker of the pair 8/2d shows relatively large upregulation in Adenomas vs. Normals (ranked 4th in the list of 67 markers), tissue 11/10a is not distinguished based on its over-expression alone (ranked 63rd of 67 markers). Further, not all highly over-expressed markers demonstrate strong discriminatory power when analyzed by cluster analysis.

Discussion of tissue classes

All clustering analysis experiments discussed here were performed using the complete set of 71 tissues available. As previously stated, no set of Markers is able to correctly classify all Normal tissues (n=21) and all Adenoma tissues (n=51). In fact, two particular Normal tissues, A7 and C5, are frequently missed using the k=3 nearest neighbor rule. The frequency of tissues missed for all 3 and 4 marker sets that achieve $\geq$ 69/71 correct classification is shown in Figure 3.

This finding demonstrates that, in terms of clustering analysis, these two Tissues yield an expression profile that is more typical of adenoma tissues than for the other healthy tissues in this pool. One hypothesis to support this result is that these two tissues are transcribing

an adenoma gene profile that proceeds, or is independent of, the morphological changes evident to the examining pathologist by histological analysis. Current theories of colorectal carcinogenesis suggest that the adenoma-carcinoma cycle is proceeded by field defects, precursors to dysplasia. While field defects have been associated with biochemical changes such as altered enzyme levels involved in proliferation, mucosal tissue sections often show no gross or histopathological changes (Young, G., Rozen, P. and Levin, B., Early Detection and Screening for Colorectal Cancer. (1996.) Saunders, New York). Therefore, it is possible that these tissues are, in fact, not normal but rather represent early adenoma development undetected by previous diagnostic techniques.

(ii) *Cluster Results by Vector Size*

Sets of 2 Markers

Using a brute-force analysis of all 2,211 possible 2-marker combinations, the highest score achieved was 68/71. This score was achieved with exactly 3 unique sets of Markers: (8/2d – 11/10a; 12/2f – 11/10a; 12/2f – 3/16bC4).

Sets of 3 markers

Analysis of all 47,905 possible 3 marker combinations yields exactly one set of markers able to identify the near perfect score of (70/71.) (See Table 9.) Exactly 6 and 60 sets of three markers were also able to correctly classify 69/71 and 68/71 tissues, respectively. Subsets yielding 69/71 are shown in Table 11.

Sets of 4 Markers

Analysis of the 766,480 possible four marker combinations yields exactly six unique subsets of markers able to correctly classify (70/71) tissues. (See Table 9.) 108 unique sets of four markers were able to correctly classify (69/71) of the tissues. (See Table 13.)

Sets of 5 or greater Markers

As the problem space for all possible five marker combinations for 67 Markers approaches 10 million, a brute-force analysis of all combinations is not practical. To analyze sets of Markers greater than five elements, a genetic algorithm was used to search the n-dimensional expression landscape for optimum or near optimum Marker sets.

Using this technique Marker sets of 5, 8, 12 and 15 Markers were explored by cluster analysis. In data collected, a number of near perfect (70/71) solution sets have been identified, however no combination has been identified that is able to perfectly classify all 71 markers.

EXAMPLE 6
SUMMARY OF PARTIALLY OR FULLY CHARACTERISED ADENOMA MARKERS

S100 Calcium binding protein P (S100P) (Seq Id Nos: 62 and 63)

Genbank accession: NM005980
Chromosome location: 4p16
Original citation:
Becker T, Gerke V, Kube E and Weber K. S100P, a novel Ca(2+)-binding protein from human placenta. cDNA cloning, recombinant protein expression and Ca2+ binding properties. Eur. J. Biochem. **207** (2), 541-547 (1992).
Transcript size: 439bp
Annotated view: Figure 4
Genbank protein accession: NP005971
Protein size: 95aa
Description:
The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs. S100 proteins are localized in the cytoplasm and/or

nucleus of a wide range of cells, and are involved in the regulation of cellular processes such as cell cycle progression and differentiation. S100 genes include at least 13 members which are located as a cluster on chromosome 1q21. However, this gene is located at 4p16. This protein, in addition to binding Ca2+, also binds Zn2+ and Mg2+. S100p is up-regulated in inflammatory diseases of the bowel such as Crohn's disease and ulcerative colitis and overexpression has been linked to breast and prostate cancer progression using model systems

Defensin α-6 (Paneth cell-specific) (Seq Id No: 49 and 50)

Genbank accession: NM001926
Chromosome location: 8pter-8p21
Original citation:
Jones DE and Bevins CL. Defensin-6 mRNA in human Paneth cells: implications for antimicrobial peptides in host defense of the human bowel. FEBS Lett. **315** (2), 187-192 (1993).
Annotated view: Figure 5
Transcript size: 475bp
Genbank protein accession: NP001917
Protein size: 100aa
Description:
Defensins are a family of antimicrobial and cytotoxic peptides thought to be involved in host defense. They are abundant in phagocytic cells of haemopoietic origin and two forms (defensin α-5 and defensin α-6) are found in the secretory granules of Paneth cells in the small intestine. The genes for the haemopoietic and enteric defensins are located in the same region of chromosome 8. The finding of an abundant defensin α-6 mRNA in human Paneth cells supports the notion that these epithelial cells may play a key role in peptide-based host defense of the bowel. Defensin α-6 is overexpressed in Crohn's disease.

Gastric intrinsic factor (GIF) (Seq Id No 43 and 44)

Genbank accession: NM005142
Chromosome location: 11q13
Original citation:
Hewitt JE, Gordon MM, Taggart RT, Mohandas TK and Alpers DH. Human gastric intrinsic factor: characterization of cDNA and genomic clones and localization to human chromosome 11. Genomics **10** (2), 432-440 (1991)
Annotated view: Figure 6
Transcript size: 1584bp
Genbank protein accession: NP005133
Protein size: 417aa
Description:
Gastric intrinsic factor (GIF) is a glycoprotein secreted by parietal cells of the gastric mucosa. GIF mediates transmembrane transport of Vitamin B_{12} via receptors that function as oligomers in the plasma membrane. GIF-mediated import of Vitamin B_{12} is limited to the apical membranes of epithelial cells. Deficiency of GIF results in pernicious anaemia.

Regenerating protein IV (RegIV) (Seq Id No: 38 and 39)

Genbank accession: NM032044
Chromosome location: 1q12-q21
Original citation:
Hartupee JC, Zhang H, Bonaldo MF, Soares MB and Dieckgraefe BK. Isolation and characterization of a cDNA encoding a novel member of the human regenerating protein family: Reg IV(1) Biochem. Biophys. Acta **1518** (3), 287-293 (2001).
Annotated view: Figure 7
Transcript size: 1200bp
Genbank protein accession: NP114433
Protein size: 158aa
Description:

Reg and Reg-related genes constitute a multi-gene family belonging to the calcium (C-type) dependent lectin superfamily. Regenerating gene family members are expressed in the proximal gastrointestinal tract and ectopically at other sites in the setting of tissue injury. Reg IV has a highly restricted tissue expression pattern, with prominent expression in the gastrointestinal tract. Reg IV mRNA expression is significantly up-regulated by mucosal injury from active Crohn's disease or ulcerative colitis. Members of the Reg gene family are known to be up-regulated in colon carcinogenesis.

GW112 protein (Seq Id No: 60 and 61)

Genbank accession: NM006418
Chromosome location: 13q14.2
Original citation:
Shinozaki S, Nakamura T, Iimura M, Kato Y, Iizuka B, Kobayashi M. and Hayashi N. Upregulation of Reg 1alpha and GW112 in the epithelium of inflamed colonic mucosa Gut **48** (5), 623-629 (2001).
Annotated view: Figure 8
Transcript size: 2840bp
Genbank protein accession: NP006409
Protein size: 218aa
Description:
This gene was originally cloned from human myeloblasts and currently its function is unknown. GW112 is selectively expressed in inflamed colonic epithelium.

Claudin-2 (Seq Id No: 2 and3)

Level of overexpression in adenomas: 250-fold
Genbank accession: NM020384
Chromosome location: Xq22.3-23
Annotated view: Figure 9
Transcript size: 2782bp

Genbank protein accession: NP065117
Protein size: 230aa
Description:
The claudins are a large family of transmembrane proteins that are part of the tight junction complex and they regulate epithelial barriers by forming structural components of a paracellular pore. Claudin-2 is found in the tight junctions of kidney, liver and intestine and is also involved in maintaining the blood-CSF barrier. The claudin-2 gene contains binding sites for and can be regulated by intestinal specific Cdx homeodomain proteins.

Solute carrier family 12, member 2 (SLC12A2) (Seq Id No: 66 and 67)

Level of overexpression in adenomas: 30-fold
Genbank accession: NM001046
Chromosome location: 5q23.3
Original citation:
Payne JA, Xu JC, Haas M, Lytle CY, Ward D and Forbush B. Primary structure, functional expression, and chromosomal localization of the bumetanide-sensitive Na-K-Cl cotransporter in human colon. J. Biol. Chem. **270** (30), 17977-17985 (1995).
Annotated view: Figure 10
Transcript size: 4375bp
Genbank protein accession: NP001037
Protein size: 1212aa
Description:
Members of the solute carrier family are Na-K-Cl cotransporters and are important for the maintenance of water and electrolyte homeostasis and aid trans-cellular movement of sodium, potassium and chloride ions in both secretory and absorptive epithelia. Expression has been observed in the thick ascending limb of the Loop of Henle in mammalian kidney and a diverse array of secretory epithelia including the intestine. It is known to exist in two forms, with polarized membrane distribution, being exclusively basolateral in distributuion within secretory epithelia, while in absorptive epithelia it is observed to localise to the

apical membrane. It should be noted that the above reference denotes the characterisation of the basolateral isoform of the cotransporter from a human colonic cell line.

Transforming growth factor, beta induced (TGFBI) (Seq Id No: 56 and 57)

Level of overexpression in adenomas: 10-fold
Genbank accession: NM000358
Chromosome location: 5q31
Original citation:
Skonier J, Neubauer M, Madisen L, Bennett K, Plowman GD and Purchio AF. cDNA cloning and sequence analysis of beta ig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-beta. DNA Cell Biol. **11** (7), 511-522 (1992).
Annotated view: Figure 11
Transcript size: 2691bp
Genbank protein accession: NP000349
Protein size: 683aa
Description:
Transforming growth factor, beta induced (TGFBI) is a protein which is induced in many cell types by TGF-β1 and is probably involved in mediating some of the signals of this growth modulator. TGFBI contains an amino-terminal secretory sequence and a ligand recognition site for several integrins at the carboxy-terminus. The TGFBI gene is located on chromosome 5q31, a region frequently deleted in preleukemic myelodysplasia and leukemia. The retinoblastoma gene protein (RB1) is known to negatively regulate the TGFBI gene as a mechanism to suppress cell growth. TGFBI has previously been shown to be up-regulated ≥ 20-fold in adenomatous and cancerous colonic epithelium. Six autosomal dominant corneal dystrophies are also caused by mutations in the TGFBI gene.

Transposon L1.1 (Seq Id No: 32, 33 and34)

Level of overexpression in adenomas: 10-fold
Genbank accession: M80340
Chromosome location: not applicable
Original citation:
Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr. Isolation of an active human transposable element. Science **254**(5039), 1805-1808 (1991).
Annotated view: Figure 12
Transcript size: 6075bp
Genbank protein accession: CDS1: AAA51621; CDS2: AAA51622
Protein size: CDS1: 41aa CDS2: 1275aa
Description:
L1 elements are retrotransposons and number from 20,000 to 50,000 in mammalian genomes making them a major component of highly repetitive DNA. A copy of the L1 element is made by the cellular RNA polymerase and is converted into double stranded DNA by a reverse transcriptase gene contained within the transposon. This copy is then inserted elsewhere in the genome. Transposable elements are thought to play an important role in evolution by creating new mutations and gene combinations and provide a mechanism to rapidly reorganise the genome. L1 elements may also be responsible for abnormal DNA rearrangement leading to carcinogenesis.

Overlapping Transcripts

The following sets of clones have been identified to overlap, at the nucleotide level, in part:

1-1d and 1-1g
11-20e and 3-19e
2-12f and 3-13e
8-12b and 5-12a

11-11b and 5-13E and 2-13B

2-17H and 2-17I-5

3-5C-3 and 3-19J-1 and 5-18F-2 and 5-19I

Without limiting the present invention in any way, these overlapping nucleic acids may represent overlapping portions of the same transcript or they may represent the existence of multiple alternative transcripts, such as splice variants.

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more of said steps or features.

Table 3

Markers with >4.5 fold upregulation

Rank	Clone Name	Fold Upregulation
1	8-7bi	296
2	11-10e	246
3	11-5b	141
4	8-2d	50
5	12-2f	45
6	4-14b	33
7	6-12a	30
8	1-6aii	29
9	2-1c	27
10	7-13b	21
11	12-7c	20
12	5-14j	17
13	5-4a	15
14	11-10b	12
15	8-19a	11
16	6-12b	10
17	5-13d	9
18	9-8a	9
19	4-11e	8
20	9-2d	8
21	11-2d	8
22	3-2c	7
23	3-12eclone3	6
24	4-2a	6
25	2-13aclone5	5
26	1-1g	5
27	6-10d	5
28	7-13dclone4	5

Table 4

Markers with 1.5-4.5 fold upregulation

Rank	Clone Name	Fold Upregulation
29	9-13c3	4
30	8-12b	4
31	humregA	4
32	9-4gclone5	4
33	2-12eclone8	4
34	7-12a	4
35	2-7gclone4	3
36	1-1d	3
37	9-8g	3
38	6-16a	3
39	11-20e	3
40	3-19e	3
41	2-1g	3
42	4-18d	3
43	3-12a	2
44	9-8f2clone5	2
45	5-2g	2
46	3-5cclone4	2
47	3-16k	2
48	8-17a	2
49	6-18dclone7	2
50	5-2f	2
51	6-17a	2
52	2-10b	2
53	2-12f	2
54	2-5d	2

Table 5

Markers with greater than 50% tissues upregulated greater than 5-fold

Rank	Clone Name	% Disease > 5 Fold
1	11-10e	100%
2	11-5b	100%
3	8-2d	100%
4	12-2f	100%
5	8-7bi	96%
6	4-14b	94%
7	5-4a	94%
8	1-6aii	90%
9	6-12a	88%
10	7-13b	86%
11	5-14j	82%
12	2-1c	80%
13	12-7c	80%
14	3-2c	75%
15	11-10b	74%
16	5-13d	74%
17	9-2d	73%
18	9-8a	72%
19	4-11e	69%
20	6-12b	68%
21	8-19a	61%
22	11-2d	60%
23	3-12eclone3	59%
24	4-2a	52%

Table 6

Markers overexpressed > 5 fold AND with Normal tissues expressing less than 20% > 5 fold

Rank	Clone Name	Fold Upregulation	% Normals > 5 fold
1	11-10e	246	14%
2	11-5b	141	14%
3	8-2d	50	19%
4	12-2f	45	19%
5	4-14b	33	14%
6	6-12a	30	19%
7	1-6aii	29	19%
8	2-1c	27	5%
9	7-13b	21	5%
10	12-7c	20	14%
11	5-14j	17	14%
12	5-4a	15	14%
13	11-10b	12	19%
14	8-19a	11	14%
15	5-13d	9	14%
16	4-11e	8	10%
17	9-2d	8	19%
18	11-2d	8	19%
19	3-2c	7	19%
20	3-12eclone3	6	0%
21	4-2a	6	19%
22	2-13aclone5	5	10%
23	1-1g	5	10%
24	6-10d	5	5%
25	7-13dclone4	5	10%

Table 7

Markers over-expressed > 5 fold in at least 50% of Adenoma tissues AND less than 20% of all Normal tissues

Rank	Clone Name	% Disease > 5 fold	% Normals > 5 fold
1	11-10e	100%	14%
2	11-5b	100%	14%
3	8-2d	100%	19%
4	12-2f	100%	19%
5	4-14b	94%	14%
6	5-4a	94%	14%
7	1-6aii	90%	19%
8	6-12a	88%	19%
9	7-13b	86%	5%
10	5-14j	82%	14%
11	2-1c	80%	5%
12	12-7c	80%	14%
13	3-2c	75%	19%
14	5-13d	74%	14%
15	11-10b	74%	19%
16	9-2d	73%	19%
17	4-11e	69%	10%
18	8-19a	61%	14%
19	11-2d	60%	19%
20	3-12eclone3	59%	0%
21	4-2a	52%	19%

Table 9

Marker sets able to classify 70/71 tissues in groups of 3 & 4.

Rank	Clone Name			
1	8-2d	4-14b	4-18e	
2	8-2d	4-14b	6-18dclone7	6-16a
3	8-2d	4-14b	6-18dclone7	5-2g
4	8-2d	4-14b	3-12eclone3	11-10a
5	8-2d	4-14b	4-18e	5-2g
6	8-2d	4-14b	4-18e	2-12f
7	12-2f	4-14b	6-16a	11-10a

Table 10

Markers useful for classifying 70/71 in groups of 3 & 4

(rank)	Clone Name
1	8-2d (12-2f)
2	4-14b
3	4-18e
4	6-18d Clone 7
5	6-16a
6	5-2g
7	11-10a
8	3-12e Clone 3 (2-12f)

Table 11

Marker sets of 3 able to classify 69/71 tissues

Rank	Clone Name		
1	12-2f	9-13c3	3-10eclone6
2	8-2d	9-13c3	3-10eclone6
3	8-2d	9-13c3	8-17a
4	12-2f	4-14b	4-18e
5	8-2d	4-14b	11-10a
6	8-2d	4-14b	6-18dclone7

5

Table 12

Markers Useful for classifying 69/71 tissues in groups of 3

(Rank)	Clone Name
9	9-13c3
10	3-10e Clone 6
11	8-17a

Table 13

Marker sets of 4 able to classify 69/71 tissues

Rank	Clone Name			
1	8-7bi	8-2d	4-14b	11-10a
2	8-2d	1-6aii	4-14b	6-18dclone7
3	8-2d	1-6aii	4-18d	11-10a
4	8-2d	12-2f	4-14b	6-18dclone7
5	8-2d	12-2f	4-14b	9-8jsclone4
6	8-2d	12-2f	4-14b	4-18e
7	8-2d	12-2f	4-14b	11-10a
8	8-2d	12-2f	2-1c	11-10a
9	8-2d	12-2f	6-18dclone7	6-10d
10	8-2d	12-2f	6-18dclone7	3-10eclone6
11	8-2d	4-14b	2-1c	6-18dclone7
12	8-2d	4-14b	2-1c	4-18e
13	8-2d	4-14b	2-1c	11-10a
14	8-2d	4-14b	6-12b	4-18e
15	8-2d	4-14b	6-12b	11-10a
16	8-2d	4-14b	8-19a	6-18dclone7
17	8-2d	4-14b	8-19a	4-18e
18	8-2d	4-14b	8-19a	11-10a
19	8-2d	4-14b	5-13d	11-10a
20	8-2d	4-14b	4-11e	11-10a
21	8-2d	4-14b	2-13aclone5	6-18dclone7
22	8-2d	4-14b	2-13aclone5	9-8jsclone4
23	8-2d	4-14b	2-13aclone5	4-18e
24	8-2d	4-14b	2-13aclone5	11-10a
25	8-2d	4-14b	6-18dclone7	4-18e
26	8-2d	4-14b	6-18dclone7	2-12f

Rank	Clone Name			
27	8-2d	4-14b	6-18dclone7	11-10a
28	8-2d	4-14b	6-18dclone7	2-5d
29	8-2d	4-14b	1-1g	11-10a
30	8-2d	4-14b	6-16a	4-18e
31	8-2d	4-14b	6-16a	11-10a
32	8-2d	4-14b	3-12eclone3	9-8jsclone4
33	8-2d	4-14b	3-12eclone3	2-18fclone5
34	8-2d	4-14b	7-13dclone4	11-10a
35	8-2d	4-14b	9-4gclone5	11-10a
36	8-2d	4-14b	8-12b	4-18e
37	8-2d	4-14b	8-12b	11-10a
38	8-2d	4-14b	2-7gclone4	11-10a
39	8-2d	4-14b	4-18e	2-5d
40	8-2d	4-14b	4-18e	8-16b
41	8-2d	4-14b	9-8g	11-10a
42	8-2d	4-14b	5-2g	11-10a
43	8-2d	4-14b	11-20e	11-10a
44	8-2d	4-14b	6-16cclone1	11-10a
45	8-2d	4-14b	2-12f	11-10a
46	8-2d	8-19a	6-18dclone7	1-1d
47	8-2d	8-19a	6-16a	3-10eclone6
48	8-2d	3-2c	6-10d	3-16bclone4
49	8-2d	2-13aclone5	7-13dclone4	3-10eclone6
50	8-2d	6-18dclone7	8-12b	3-10eclone6
51	8-2d	6-10d	3-12eclone3	3-10eclone6
52	8-2d	6-10d	7-13dclone4	3-10eclone6
53	8-2d	6-10d	8-12b	3-16bclone4
54	8-2d	6-10d	9-8g	3-16bclone4
55	8-2d	6-10d	3-10eclone6	2-5d

Rank	Clone Name			
56	8-2d	3-12eclone3	9-13c3	3-10eclone6
57	8-2d	9-13c3	8-17a	2-12f
58	8-2d	9-13c3	9-8g	3-10eclone6
59	8-2d	9-13c3	2-12f	3-10eclone6
60	8-2d	8-12b	2-12f	3-10eclone6
61	1-6aii	12-2f	4-14b	6-18dclone7
62	1-6aii	12-2f	4-14b	4-18e
63	1-6aii	12-2f	4-14b	11-10a
64	1-6aii	12-2f	8-19a	6-18dclone7
65	1-6aii	12-2f	9-8g	3-10eclone6
66	1-6aii	12-2f	4-18d	11-10a
67	12-2f	4-14b	2-1c	6-18dclone7
68	12-2f	4-14b	2-1c	6-16cclone1
69	12-2f	4-14b	2-1c	11-10a
70	12-2f	4-14b	5-13d	11-10a
71	12-2f	4-14b	4-11e	11-10a
72	12-2f	4-14b	2-13aclone5	11-10a
73	12-2f	4-14b	6-18dclone7	5-2g
74	12-2f	4-14b	6-16a	9-8jsclone4
75	12-2f	4-14b	3-12eclone3	2-18fclone5
76	12-2f	4-14b	3-12eclone3	11-10a
77	12-2f	4-14b	9-13c3	11-10a
78	12-2f	4-14b	8-12b	4-18e
79	12-2f	4-14b	4-18e	5-2g
80	12-2f	4-14b	5-2g	11-10a
81	12-2f	4-14b	11-10a	8-16b
82	12-2f	2-1c	6-10d	3-10eclone6
83	12-2f	11-10b	4-11e	3-16bclone4
84	12-2f	9-2d	11-2d	3-10eclone6

Rank	Clone Name			
85	12-2f	3-2c	4-11e	3-16bclone4
86	12-2f	5-13d	6-10d	3-10eclone6
87	12-2f	5-13d	3-12eclone3	12-17a
88	12-2f	2-13aclone5	6-10d	4-18e
89	12-2f	2-13aclone5	6-10d	3-16bclone4
90	12-2f	2-13aclone5	7-13dclone4	3-10eclone6
91	12-2f	6-18dclone7	9-13c3	3-16bclone4
92	12-2f	6-18dclone7	9-13c3	2-12f
93	12-2f	6-18dclone7	9-13c3	8-16b
94	12-2f	6-10d	9-13c3	4-18e
95	12-2f	6-10d	7-13dclone4	3-10eclone6
96	12-2f	6-10d	8-12b	3-10eclone6
97	12-2f	6-10d	5-2g	3-10eclone6
98	12-2f	6-10d	3-16bclone4	3-10eclone6
99	12-2f	3-12eclone3	9-13c3	3-10eclone6
100	12-2f	3-12eclone3	4-17d	3-10eclone6
101	12-2f	9-13c3	7-13dclone4	3-10eclone6
102	12-2f	9-13c3	8-12b	3-10eclone6
103	12-2f	9-13c3	9-8g	3-10eclone6
104	12-2f	9-13c3	5-2g	3-10eclone6
105	12-2f	9-13c3	2-12f	3-10eclone6
106	12-2f	9-13c3	2-12f	11-10a
107	12-2f	7-13dclone4	4-18e	9-8g
108	12-2f	7-13dclone4	9-8g	3-10eclone6

Table 14

Markers Useful for classifying 69/71 tissues in groups of 4

(rank)	Clone Name
12	6-10d
13	2-13a Clone 5
14	1-6a ii
15	2-1c
16	7-13d Clone 4
17	8-12b
18	3-16b Clone 4
19	9-8g
20	8-19a
21	5-13d (4-11e)
22	9-8js Clone4
23	2-5d
24	8-16b
25	6-12b
26	3-2c (11-10b, 9-2d, 11-2d)
27	4-18d
28	2-18f Clone 5
29	6-16c Clone 1
30	8-7 bi
31	1-1g
32	9-4g Clone5
33	2-7g Clone 4
34	1-1d
35	11-20e
36	4-17d
37	12-17a

Table 15

Marker sets of 3 that provide perfect classification of all tissues (69/69)

Rank	Clone Name		
1	8-2d	5-13d	3-10eClone6
2	8-2d	4-11e	3-10eClone6
3	8-2d	6-10d	3-10eClone6
4	8-2d	9-13c3	3-10eClone6
5	12-2f	5-13d	3-10eClone6
6	12-2f	4-11e	3-10eClone6
7	12-2f	2-13aclone5	3-10eClone6
8	12-2f	4-14b	3-10eClone6

BIBLIOGRAPHY

Alon, A., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mach, D. and Levine, A.J. *Proc. Natl. Acad. Sci. USA: 96*, 6745-6750, June 1999

Becker T, Gerke V, Kube E and Weber K. *Eur. J. Biochem. 207 (2)*, 541-547 (1992).

Bonner *et al* (1973) *J. Mol. Biol. 81:*123

Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr. *Science 254(5039)*, 1805-1808 (1991).

Douillard and Hoffman, Basic Facts about Hybridomas, in *Compendium of Immunology* Vol II, ed. by Schwartz, 1981;

Hartupee JC, Zhang H, Bonaldo MF, Soares MB and Dieckgraefe BK. *Biochem. Biophys. Acta 1518 (3)*, 287-293 (2001).

Hewitt JE, Gordon MM, Taggart RT, Mohandas TK and Alpers DH. *Genomics 10 (2)*, 432-440 (1991)

Kohler and Milstein, *Nature 256*: 495-499, 1975

Jones DE and Bevins CL. *FEBS Lett. 315 (2)*, 187-192 (1993).

Kohonen, T. (1995) Self-Organizing Maps. Springer Verlad, Berlin. 17-18.

Li, L., Darden, T., Weinberg, C., Levine, A. and Pedersen, L. (2001) Gene Assessment and Sample classification for gene expression data using a genetic algorithm/k nearest neighbor method. Combinatorial Chemistry & High Throughput Screening, Vol. 4(8), 727-739.

Moore, A., Basilion, J., Chiocca, e., and Weissleder, R., Measuring Transferrin Receptor Gene Expression by NMR Imaging. BBA, 1402:239-249, 1988

Payne JA, Xu JC, Haas M, Lytle CY, Ward D and Forbush B. *J. Biol. Chem. 270 (30)*, 17977-17985 (1995).

Shinozaki S, Nakamura T, Iimura M, Kato Y, Iizuka B, Kobayashi M. and Hayashi N. Upregulation of Reg 1alpha and GW112 in the epithelium of inflamed colonic mucosa Gut **48** (5), 623-629 (2001).

Skonier J, Neubauer M, Madisen L, Bennett K, Plowman GD and Purchio AF. *DNA Cell Biol. 11 (7)*, 511-522 (1992).

Wedemeyer, N., Potter, T., Wetzlich, S. and Gohde, W. Flow Cytometric Quantification of Competitive Reverse Transcriptase-PCR products, Clinical Chemistry 48:9 1398-1405, 2002

Weissleder, R., Moore, A., Ph.D., Mahmood-Bhorade, U., Benveniste, H., Chiocca, E.A., Basilion, J.P. High resolution *in vivo* imaging of transgene expression, Nature Medicine, 6:351-355, 2000

Young, G., Rozen, P. and Levin, B., Early Detection and Screening for Colorectal Cancer. (1996.) Saunders, New York.

Young, G.P., Rozen, P. and Levin, B. Chapter 3: How Does Colorectal Cancer Develop? Colorectal cancer in Clinical Practice, Ed. Rozen, P., Young, G.P., Levin, P., Spann, S.J. Martin Dunitz 2002